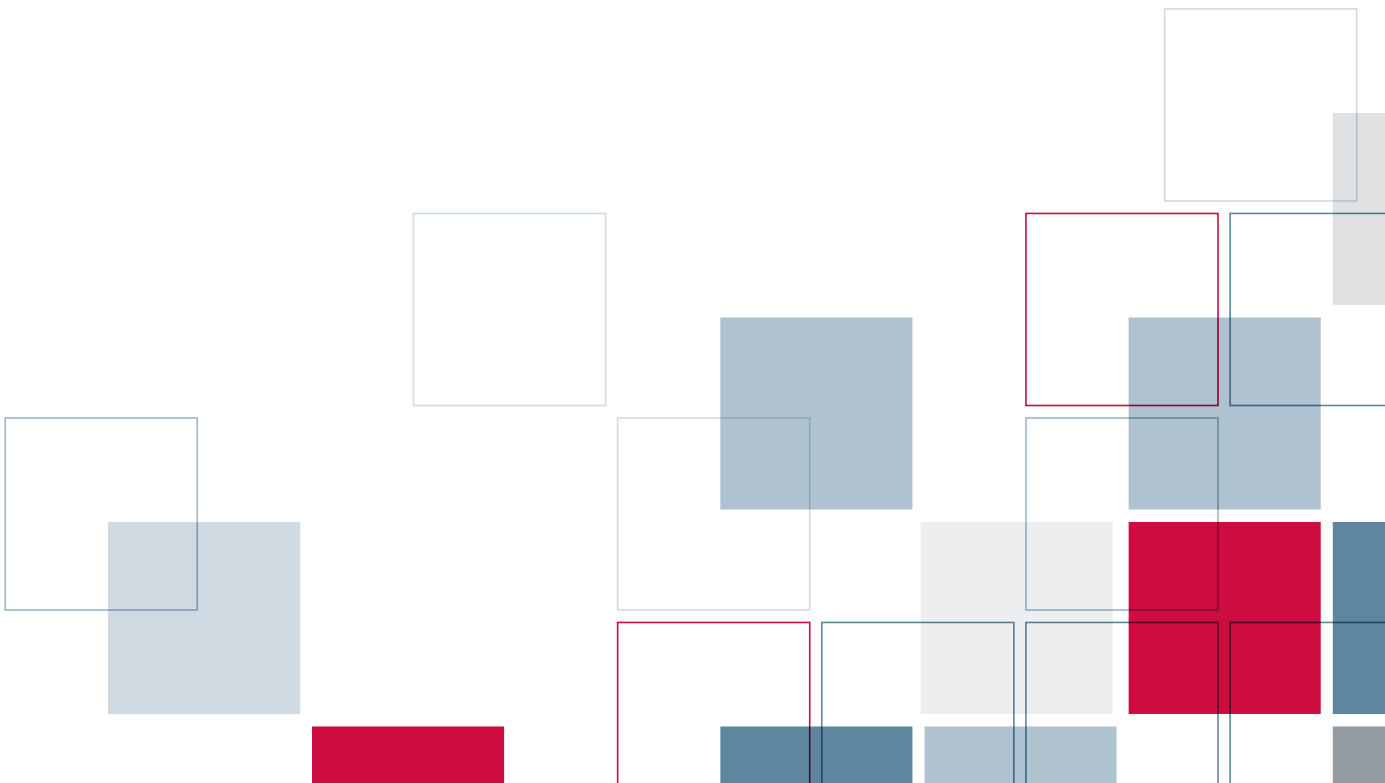
SPSS® Base 14.0
Manuale dell'utente



Per ulteriori informazioni sui prodotti software SPSS®, visitare il sito Web all'indirizzo *http://www.spss.it* o contattare:

SPSS Italia
Via Isonzo 61
Centro Direzionale Meridiana
Casalecchio di Reno
40033 BOLOGNA
Tel 800 437300
Fax 800 144160

SPSS è un marchio registrato e gli altri nomi di prodotti sono marchi di software di proprietà di SPSS Inc. Non è consentito produrre o distribuire materiale che descriva il software senza previa autorizzazione scritta dei proprietari dei diritti di marchio e dei diritti di licenza per il software, nonché dei diritti di copyright per la documentazione.

Nota: Altri nomi di prodotti menzionati nel presente documento vengono utilizzati esclusivamente a fini identificativi e sono marchi dei rispettivi proprietari.

TableLook è un marchio di fabbrica di SPSS Inc.
Windows è un marchio registrato di Microsoft Corporation.
DataDirect, DataDirect Connect, INTERSOLV e SequeLink sono marchi registrati di DataDirect Technologies.
Alcune parti di questo prodotto sono state create con LEADTOOLS © 1991–2000, LEAD Technologies, Inc. TUTTI I DIRITTI RISERVATI.
LEAD, LEADTOOLS e LEADVIEW sono marchi registrati di LEAD Technologies, Inc.
Sax Basic è un marchio di fabbrica di Sax Software Corporation. Copyright © 1993–2004 by Polar Engineering and Consulting. Tutti i diritti riservati.
Acune parti del prodotto sono state realizzate con la collaborazione di FreeType Team (*http://www.freetype.org*).
Una parte del software SPSS utilizza la tecnologia zlib. Copyright © 1995–2002 Jean-loup Gailly e Mark Adler. Il software zlib viene fornito "come tale" senza alcuna garanzia espressa o implicita.
Alcune parti del software SPSS contengono librerie Sun Java Runtime. Copyright © 2003 Sun Microsystems, Inc. Tutti i diritti riservati. Le librerie run-time Sun Java includono codice concesso in licenza da RSA Security, Inc. Alcune parti delle librerie sono concesse in licenza da IBM e sono disponibili all'indirizzo *http://oss.software.ibm.com/icu4j/*.

SPSS Base 14.0 Manuale dell'utente
Copyright © 2005 SPSS Inc.
Tutti i diritti riservati.

Nessuna parte di questa pubblicazione potrà comunque essere riprodotta o inserita in un sistema di riproduzione o trasmessa in qualsiasi forma e con qualsiasi mezzo (in formato elettronico, meccanico, su fotocopia, come registrazione o altro) per qualsiasi scopo, senza il permesso scritto dell'editore.

1 2 3 4 5 6 7 8 9 0 08 07 06 05

ISBN 1-56827-681-8

Prefazione

SPSS 14.0

SPSS 14.0 è un sistema completo per l'analisi dei dati. SPSS è in grado di elaborare dati provenienti da quasi tutti i tipi di file e di utilizzarli per generare rapporti con tabelle, grafici, grafici di distribuzioni e trend, statistiche descrittive e complesse analisi statistiche.

Questo manuale fornisce *SPSS Base 14.0 Manuale dell'utente* indicazioni sull'interfaccia grafica di SPSS per Windows. Nell'Aiuto in linea installato con il software sono forniti esempi relativi all'utilizzo delle procedure statistiche disponibili in SPSS Base 14.0. Gli algoritmi usati nelle procedure statistiche vengono forniti in un documento PDF ma possono anche essere visualizzati dal menu ?.

Oltre ai menu e alle finestre di dialogo, in SPSS è possibile utilizzare un linguaggio a comandi. Alcune caratteristiche avanzate del sistema possono essere utilizzate soltanto mediante la sintassi dei comandi. Tali caratteristiche non sono disponibili nella versione per studenti. Informazioni dettagliate sui riferimenti alla sintassi dei comandi sono disponibili in due formati: nella Guida integrate e su un documento PDF a parte in *SPSS 14.0 Command Syntax Reference*, anch'esso consultabile tramite il menu ?.

Moduli opzionali di SPSS

Le seguenti opzioni sono disponibili come componenti avanzati aggiuntivi per il sistema completo SPSS Base (non per la versione per studenti):

SPSS Regression Models™ fornisce tecniche per l'analisi dei dati che non rientrano nei modelli statistici lineari tradizionali. Include procedure per l'analisi Probit, la regressione logistica, il metodo dei minimi quadrati ponderati, la regressione con il metodo dei minimi quadrati ponderati e la regressione non lineare generalizzata.

SPSS Advanced Models™ offre tecniche spesso utilizzate nella ricerca biomedica e sperimentale avanzata. Include procedure per modelli lineari generalizzati (GLM), modelli misti lineari, analisi della varianza, analisi loglineare, regressione ordinale, tavole di sopravvivenza attuariali, analisi della sopravvivenza di Kaplan-Meier e regressione di Cox di base ed estesa.

SPSS Tables™ consente di creare una vasta gamma di rapporti con tabelle, incluse complesse tabelle con intestazioni e colonne, e consente di visualizzare dati a risposta multipla.

SPSS Trends™ consente di eseguire previsioni complete e analisi di serie storiche con più modelli per la stima delle curve, modelli di livellamento e metodi per la stima delle funzioni di autoregressione.

SPSS Categories® consente di eseguire procedure di scaling ottimale, inclusa l'analisi delle corrispondenze.

SPSS Conjoint™ consente di eseguire l'analisi congiunta.

SPSS Exact Tests™ consente di calcolare valori *p* esatti nei test statistici, soprattutto quando, a causa di campioni piccoli o mal distribuiti, i test comuni potrebbero risultare non corretti.

SPSS Missing Value Analysis™ descrive i dati mancanti, fornisce una stima delle medie e di altre statistiche e assegna i valori per le osservazioni mancanti.

SPSS Maps™ consente di trasformare dati geografici in mappe di alta qualità con simboli, colori, grafici a barre e combinazioni di tematiche, per rappresentare graficamente non solo cosa accade, ma anche dove accade.

SPSS Complex Samples™ consente ai ricercatori che eseguono sondaggi, indagini di mercato, ricerche in ambito medico o dell'opinione pubblica, nonché agli scienziati che si avvalgono della metodologia basata sui sondaggi campione, di incorporare nell'analisi dei dati i disegni di campioni complessi creati.

La procedura **Albero di classificazione di SPSS™** crea un modello di classificazione basato su alberi. Essa classifica i casi in gruppi o prevede i valori di una variabile dipendente (di destinazione) in base ai valori di variabili (predittore) indipendenti. La procedura offre strumenti di validazione per l'analisi di classificazione confermativa ed esplorativa.

La procedura **Convalida dei dati di SPSS ™** fornisce un'istantanea immediata dei dati. In particolare consente di applicare le regola di convalida che identificano i valori di dati non validi. È possibile creare regole che escludono i valori fuori intervallo, i

valori mancanti o vuoti. È possibile anche salvare le variabili che registrano violazioni alle singole regole nonché il numero totale di violazioni alle regole per caso. Viene fornito un set limitato di regole predefinite da copiare o modificare.

La famiglia di prodotti SPSS include inoltre applicazioni per l'immissione di dati, analisi del testo, classificazione, reti e creazione di diagrammi di flusso.

Installazione

Per installare SPSS Base system, eseguire l'Attivazione guidata licenza utilizzando il codice di autorizzazione fornito da SPSS Inc. Per ulteriori informazioni, consultare le istruzioni di installazione fornite con il SPSS Base system.

Compatibilità

SPSS è stato progettato per essere eseguito con numerosi sistemi operativi. Per informazioni specifiche sui requisiti minimi e consigliati, vedere le istruzioni di installazione fornite con il sistema.

Numeri di serie

Il numero di serie è il numero di identificazione del cliente per SPSS Inc. Sarà necessario fornire questo numero nel caso l'utente contatti SPSS Inc. per ricevere informazioni relative al supporto, al pagamento e ai sistemi aggiornati. Il numero di serie viene fornito con il modulo Base.

Servizio clienti

Per informazioni sulla spedizione o sul proprio account, rivolgersi alla filiale SPSS del proprio paese, ai recapiti riportati sul sito Web di SPSS all'indirizzo *http://www.spss.com/worldwide*. Tenere presente che sarà necessario fornire il numero di serie.

Corsi di formazione

SPSS Inc. organizza corsi di formazione su SPSS, che includono esercitazioni pratiche. Tali corsi si terranno periodicamente nelle principali città. Per ulteriori informazioni sui corsi, contattare la filiale SPSS nel proprio paese, indicata nel sito Web SPSS all'indirizzo *http://www.spss.com/worldwide*.

Supporto tecnico

Ai clienti registrati SPSS mette a disposizione un servizio di supporto tecnico. I clienti possono contattare il supporto tecnico per richiedere assistenza sull'utilizzo dei prodotti SPSS o sull'installazione di uno degli ambienti hardware supportati. Per il supporto tecnico, visitare il sito Web di SPSS all'indirizzo *http://www.spss.it* o contattare la filiale SPSS nel proprio paese indicata nel sito Web di SPSS all'indirizzo *http://www.spss.com/worldwide*. Tenere presente che sarà necessario fornire dati di identificazione personali e relativi alla propria società e il numero di serie del sistema.

Pubblicazioni aggiuntive

È possibile acquistare i manuali relativi ai prodotti SPSS direttamente presso SPSS Inc., visitando il SPSS Web Store all'indirizzo *http://www.spss.com/estore* oppure contattando la filiale SPSS del proprio paese ai recapiti riportati sul sito Web di SPSS all'indirizzo *http://www.spss.com/worldwide*. Per effettuare ordini telefonicamente negli Stati Uniti e in Canada, chiamare SPSS Inc. al numero 800-543-2185. Per effettuare ordini al di fuori del Nord America, contattare la filiale SPSS nel proprio paese indicata nel sito Web di SPSS.

Presso Prentice Hall sarà possibile ordinare a breve il volume *SPSS Statistical Procedures Companion* di Marija Norušis, che sarà presto disponibile in versione aggiornata per SPSS 14.0. Verrà presto pubblicato anche il volume *SPSS Advanced Statistical Procedures Companion*, basato su SPSS 14.0. È in fase di preparazione il volume *SPSS Guide to Data Analysis* per SPSS 14.0 è in fase di sviluppo. I titoli dei volumi disponibili esclusivamente tramite Prentice Hall verranno pubblicati sul sito Web di SPSS all'indirizzo *http://www.spss.com/estore* (selezionare il proprio paese di residenza, quindi fare clic su Books).

Scrivici

SPSS considera prezioso qualsiasi suggerimento. SPSS accoglierà con entusiasmo qualsiasi commento relativo all'utilizzo dei prodotti SPSS e tutte le informazioni sulle nuove e interessanti applicazioni ottenute mediante SPSS Base system. Inviare un messaggio di posta elettronica a *suggest@spss.com* oppure scrivere a SPSS Inc., Attn.: Director of Product Planning, 233 South Wacker Drive, 11th Floor, Chicago, IL 60606-6412.

Informazioni sul manuale

In questo manuale viene descritta e documentata l'interfaccia utente grafica relativa alle procedure disponibilinel modulo SPSS Base system. Le immagini delle finestre di dialogo provengono da SPSS for Windows. Le finestre di dialogo di altri sistemi operativi sono simili. Per informazioni dettagliate sulla sintassi dei comandi delle funzioni di SPSS Base system sono disponibili in due formati: nella Guida in linea integrata e in un documento PDF a parte che può essere visualizzato tramite *SPSS 14.0 Command Syntax Reference*, visualizzabile tramite il menu ?.

Contattaci

Per registrarsi alla mailing list, rivolgersi a uno dei nostri uffici ai recapiti indicati alla pagina Web *http://www.spss.com/worldwide*.

Sommario

1 Osservazioni generali 1

Novità di SPSS 14.0. 2
Finestre 6
Menu 9
Barra di stato 10
Finestre di dialogo 10
Nomi ed etichette di variabili negli elenchi delle finestre di dialogo. 11
Controlli delle finestre di dialogo. 12
Sottofinestre di dialogo. 12
Selezione di variabili. 13
Icone dell'elenco Variabili. 13
Per ottenere informazioni sulle variabili in una finestra di dialogo 13
Passaggi fondamentali dell'analisi dei dati 14
Analisi guidate 15
Ulteriori informazioni su SPSS 15

2 Aiuto 17

Utilizzo del sommario dell'Aiuto. 20
Utilizzo dell'Indice dell'Aiuto 21
Uso della funzione di ricerca della Guida 22
Visualizzazione di informazioni sui controlli delle finestre di dialogo 24
Visualizzazione di informazioni sui termini delle tabelle pivot dell'output . . . 24
Utilizzo degli studi di casi 25
Per copiare il testo dell'Aiuto da una finestra popup 25

3 File di dati 27

Apertura di un file di dati. 27
Per aprire file di dati 27
Tipi di file di dati 28
Apertura file: opzioni 29
Lettura di file di Excel 5 o versione successiva 29
Per leggere file Excel di versioni precedenti e altri formati di fogli elettronici 30
Lettura di file dBASE. 30
Lettura di file Stata 31
Lettura di file di database 31
Importazione guidata di testo 51
Lettura di dati Dimensions 62
Informazioni sul file. 68
Salvataggio di file di dati. 69
Salvataggio di file di dati modificati. 69
Salvataggio di file di dati in formato Excel. 69
Salvataggio di file di dati in formato SAS. 70
Salvataggio di file di dati in formato Stata 72
Salvataggio di un nuovo file di dati SPSS o in formato diverso. 73
Salvataggio di dati: Tipi di file di dati 74
Salvataggio di sottoinsiemi di variabili 76
Opzioni per il salvataggio di file. 77
Protezione dei dati originali. 77
File attivo virtuale 78

4 Analisi distribuita 83

Confronto tra analisi distributa e locale. 83

5 Editor dei dati 95

Visualizzazione dati. 96
Visualizzazione variabili 97
Inserimento di dati 108
Modifica dei dati nella Visualizzazione dati 111
Vai al caso 114
Criteri di selezione dei casi nell'Editor dei dati. 115
Editor dei dati: Opzioni di visualizzazione. 115
Stampa dall'Editor dei dati 117

6 Utilizzo di più sorgenti dati 119

Gestione di più sorgenti di dati 120
Copia e inserimento di informazioni in più file di dati 122
Ridenominazione dei file di dati. 122

7 Preparazione dei dati 123

Proprietà variabili 123
Definizione di proprietà delle variabili 124
Insiemi a risposta multipla 132
Copia delle proprietà dei dati 136
Identificazione di casi duplicati. 146
Segmentazione grafica 149

8 Trasformazioni di dati 161

Calcolo delle variabili 161
Funzioni 164
Valori mancanti nelle funzioni 165
Generatore di numeri casuale 165
Conteggia occorrenze di valori all'interno dei casi 167
Ricodifica dei valori 169
Ricodifica nelle stesse variabili 169
Ricodifica in variabili differenti 172
Classifica casi 175
Ricodifica automatica 179
Impostazione guidata data e ora 183
Trasformazioni di dati di serie storiche 198
Calcolo del punteggio dei dati con modelli predittivi 205

9 Gestione e trasformazione di file 211

Ordina casi 212
Trasponi 213
Unione di file di dati 213
Aggiungi casi 214
Aggiungi variabili 218
Aggrega dati 221
Distingui 226
Seleziona casi 227
Pesa casi 232
Ristrutturazione di dati 234

10 Utilizzo dell'output 257

Viewer 257
Utilizzo dell'output in altre applicazioni 265
Per incollare oggetti nel Viewer 269
Incolla speciale 269
Per incollare oggetti da altre applicazioni nel Viewer 270
Esporta output 270
Stampa di documenti del Viewer. 282
Salvataggio dell'output 289

11 Bozza 291

Per creare output della Bozza 292
Controllo del formato dell'output bozza 293
Caratteri in Bozza 298
Per stampare l'output della Bozza 298
Per salvare l'output della Bozza come testo 300

12 Tabelle pivot 301

Gestione di una tabella pivot 301
Utilizzo degli strati. 306
Segnalibri 310
Per visualizzare e nascondere le celle 311
Modifica dei risultati 313
Modifica dell'aspetto delle tabelle 314
Proprietà tabella 316

Per modificare le proprietà delle tabelle pivot 316
Proprietà tabella: Generale 316
Proprietà tabella: Note a piè di pagina 317
Proprietà tabella: Formati di cella 319
Proprietà tabella: Bordi 320
Proprietà tabella: Stampa 322
Carattere 323
Larghezza delle celle 324
Proprietà cella 325
Proprietà cella: Valore 326
Proprietà cella: Allineamento 327
Proprietà cella: Margini 328
Proprietà cella: Ombreggiatura 329
Rimando a nota 330
Selezione di righe e colonne nelle tabelle pivot 331
Per selezionare una riga o una colonna in una tabella pivot 331
Modifica dei risultati della tabella pivot 331
Stampa delle tabelle pivot 332
Per stampare gli strati nascosti di una tabella pivot 333
Impostazione delle interruzioni di pagina per le tabelle con larghezza o lunghezza eccessiva 333

13 Utilizzo della sintassi dei comandi 335

Regole della sintassi dei comandi 335
Creazione di sintassi dei comandi nelle finestre di dialogo 337
Copia della sintassi dal registro dell'output 339
Utilizzo di sintassi di un file giornale 340

Per eseguire la sintassi dei comandi 342
Comandi EXECUTE multipli 342

14 Frequenze 345

Frequenze: Statistiche 348
Frequenze: Grafici 350
Frequenze: Formato 351

15 Descrittive 353

Descrittive: Opzioni 355
Funzioni aggiuntive del comando DESCRIPTIVES 356

16 Esplora 359

Esplora: Statistica 363
Esplora: Grafici 364
Esplora: Opzioni 366
Funzioni aggiuntive del comando EXAMINE 366

17 Tavole di contingenza 369

Strati nelle tavole di contingenza 372
Tavole di contingenza: Grafici a barre raggruppati 372
Tavole di contingenza: Statistiche 372

Tavole di contingenza: Visualizzazione cella 376
Tavole di contingenza: Formato tabella 377

18 Riassumi 379

Riassumi: Opzioni 381
Riassumi: Statistiche 382

19 Medie 385

Medie: Opzioni 388

20 Cubi OLAP 391

Cubi OLAP: Statistiche 394
Cubi OLAP: Differenze. 397
Cubi OLAP: Titolo 398

21 Test T 399

Test T per campioni indipendenti. 399
Test T per campioni appaiati 403
Test T per un campione. 406
Opzioni aggiuntive del comando T-TEST 409

22 ANOVA univariata 411

ANOVA univariata: Contrasti 414
ANOVA univariata: Test Post Hoc 415
ANOVA univariata: Opzioni 418
Opzioni aggiuntive del comando ONEWAY 419

23 Analisi GLM univariato 421

GLM – Univariato: Modello 425
GLM - Univariato: Contrasti 428
GLM – Univariato: Profili 429
GLM - Univariato: Confronti post hoc 431
GLM – Univariato: Salva 433
GLM - Univariato: Opzioni 436
Funzioni aggiuntive del comando UNIANOVA 437

24 Correlazioni bivariate 439

Correlazioni bivariate: Opzioni 442
Funzioni aggiuntive dei comandi CORRELATIONS e NONPAR CORR. 443

25 Correlazioni parziali 445

Correlazioni parziali: Opzioni 448
Opzioni aggiuntive del comando PARTIAL CORR 448

26 Distanze 451

Distanze: Misure di dissimilarità 453
Distanze: Misure di similarità 454
Opzioni aggiuntive del comando PROXIMITIES 455

27 Regressione lineare 457

Metodi di selezione della variabile di regressione lineare 461
Regressione lineare: Imposta regola. 463
Regressione lineare: grafici 463
Regressione lineare: Per salvare nuove variabili. 465
Regressione lineare: Statistiche 468
Regressione lineare: Opzioni 470
Opzioni aggiuntive del comando REGRESSION 471

28 Stima di curve 473

Stima di curve: Modelli. 477
Stima di curve: Salva 478

29 Analisi discriminante 481

Analisi discriminante: Definisci intervallo 484
Analisi discriminante: Seleziona casi 484
Analisi discriminante: Statistiche 485
Analisi discriminante: Metodo Stepwise. 486

Analisi discriminante: Classificazione 488
Analisi discriminante: Salva 490
Funzioni aggiuntive del comando DISCRIMINANT. 490

30 Analisi fattoriale 493

Analisi fattoriale: Seleziona casi 497
Analisi fattoriale: Descrittive. 498
Analisi fattoriale: Estrazione 499
Analisi fattoriale: Rotazione 501
Analisi fattoriale: Punteggi fattoriali 502
Analisi fattoriale: Opzioni 503
Opzioni aggiuntive del comando FACTOR 503

31 Scelta di una procedura di cluster analysis 505

Scelta di una procedura per il raggruppamento 505

32 Analisi Cluster TwoStep 507

Opzioni di Analisi Cluster TwoStep 511
Grafici di Analisi Cluster TwoStep 514
Output di Analisi Cluster TwoStep 515

33 Cluster gerarchica 517

Cluster gerarchica: Metodo 521
Cluster gerarchica: Statistiche 522
Cluster gerarchica: Grafici 523
Cluster gerarchica: Salva nuove variabili 523
Funzioni aggiuntive della sintassi del comando CLUSTER 524

34 Cluster con metodo delle K-medie 525

Efficienza dell'analisi cluster K-medie 529
Cluster K-medie: Iterazioni 530
Cluster K-medie: Salva 531
Cluster K-medie: Opzioni. 531
Opzioni aggiuntive del comando QUICK CLUSTER 532

35 Test non parametrici 533

Test Chi-quadrato 534
Test binomiale 538
Test delle successioni. 541
Test di Kolmogorov-Smirnov per un campione. 544
Test per due campioni indipendenti. 547
Test per due campioni dipendenti 551
Test per diversi campioni indipendenti 554
Test per diversi campioni dipendenti 558

36 Analisi a risposta multipla 561

Risposte multiple: Definisci insiemi 562
Risposte multiple: Frequenze 564
Risposte multiple: Tavole di contingenza 566
Risposte multiple, tavole di contingenza: Definisci intervalli delle variabili . 568
Risposte multiple, tavole di contingenza: Opzioni 569
Funzioni aggiuntive del comando MULT RESPONSE 570

37 Report 571

Risultati di report 571
Report: Riepiloghi per righe 571
Report: Riepiloghi per colonne 580
Funzioni aggiuntive del comando REPORT 586

38 Analisi di affidabilità 587

Analisi di affidabilità: Statistiche 589
Opzioni aggiuntive del comando RELIABILITY 591

39 Scaling multidimensionale 593

Scaling multidimensionale: Forma dei dati 595
Scaling multidimensionale: Crea misure dai dati 596
Scaling multidimensionale: Modello 597
Scaling multidimensionale: Opzioni 598

Opzioni aggiuntive del comando ALSCAL 599

40 Statistiche di rapporto 601

Statistiche di rapporto 603

41 Panoramica dello strumento Grafico 605

Creazione e modifica di un grafico 605
Opzioni definizione grafico 612

42 Curva ROC 619

Curva ROC: Opzioni 622

43 Strumenti 623

Informazioni sulle variabili 623
Commenti sul file di dati 624
Insiemi di variabili. 625
Definisci insiemi di variabili. 625
Usa insiemi. 626
Riordinamento degli elenchi di variabili di destinazione. 627

44 Opzioni 629

Opzioni generali 630
Opzioni Viewer 632
Opzioni Output bozza 633
Opzioni: Etichette di output 635
Opzioni: grafici 637
Opzioni dei grafici interattivi 641
Opzioni tabella pivot 643
Opzioni: Dati 645
Opzioni: Valuta 646
Opzioni: Script 648

45 Personalizzazione dei menu e delle barre degli strumenti 651

Editor dei menu 651
Personalizzazione delle barre degli strumenti 652
Visualizza barre degli strumenti 652
Per personalizzare le barre degli strumenti 653

46 Utilità SPSS Production 659

. 659
Opzioni di esportazione 662
Prompt utente 664
Prompt macro dell'utilità SPSS Production 666
Opzioni dell'utilità SPSS Production 667

Impostazione del formato per le sessioni dell'utilità SPSS Production 668
Esecuzione di sessioni dell'utilità SPSS Production dalla riga di comando . 671
Pubblica sul Web 673
SmartViewer Web Server: Login 674

47 Utilità Script di SPSS 675

Per eseguire uno script. 675
Script disponibili in SPSS 676
Script automatici 677
Creazione e modifica di script. 679
Per modificare uno script 680
finestra dello script. 680
Script di avvio. 683
Creazione di script automatici. 684
Modalità di funzionamento degli script 689
Convenzioni di denominazione per le classi e le variabili oggetto SPSS. . . 691
Nuova procedura (script) 696
Aggiunta di descrizioni agli script 699
Script per la creazione di finestre di dialogo personalizzate 699
Debug degli script. 704
File di script e di sintassi. 707

48 Sistema di gestione dell'output 711

Tipi di oggetto di output. 715
Identificatori di comando e sottotipi di tabelle. 717
Etichette. 718

Opzioni SGO 720
Registrazione 725
Esclusione della visualizzazione dell'output in Viewer 726
Trasmissione dell'output ai file dati SPSS 726
Struttura della tabella OXML. 737
Identificatori SGO 741

Appendici

A Database Access Administrator *745*

B Personalizzazione dei documenti HTML *747*

Per aggiungere codice HTML personalizzato a documenti di output esportati 747
Contenuto e formato del file di testo per HTML personalizzato 748
Per utilizzare un file o una posizione diversa per il codice HTML personalizzato 748

Indice *751*

Capitolo

1

Osservazioni generali

SPSS per Windows è un valido sistema per l'analisi statistica e la gestione dei dati in un ambiente grafico in cui è possibile utilizzare menu descrittivi e semplici finestre di dialogo per eseguire automaticamente numerose operazioni. Con un solo clic del mouse è possibile eseguire la maggior parte delle operazioni.

Oltre all'interfaccia per l'analisi statistica, di facile utilizzo in quanto prevede l'impiego del mouse per numerose operazioni, in SPSS per Windows sono disponibili anche:

Editor dei dati. L'Editor dei dati è un sistema versatile dotato delle stesse caratteristiche di un foglio elettronico per la definizione, l'inserimento, la modifica e la visualizzazione dei dati.

Viewer. Il Viewer consente di sfogliare rapidamente i risultati, visualizzare e nascondere l'output in modo selettivo, modificare l'ordine di visualizzazione dei risultati e trasferire tabelle e grafici di alta qualità da SPSS ad altre applicazioni.

Tabelle pivot multidimensionali. È possibile mettere in risalto i risultati grazie alle tabelle pivot multidimensionali. In queste tabelle i risultati possono essere visualizzati in modi diversi modificando la disposizione di righe, colonne e strati. È possibile mettere in risalto risultati importanti che in genere non compaiono nei tipi di report più comuni. I gruppi possono essere facilmente confrontati dividendo la tabella in modo da visualizzare un solo gruppo per volta.

Grafici ad alta risoluzione. SPSS include come funzioni standard grafici a torta, grafici a barre, istogrammi, grafici a dispersione, grafici 3-D e di altro tipo, tutti a colori e ad alta risoluzione.

Accesso ai database. È possibile recuperare informazioni dai database utilizzando Creazione guidata di query anzichè query SQL complicate.

Trasformazioni di dati. Le funzioni di trasformazione risultano utili per preparare i dati alle analisi. È possibile creare sottoinsiemi di dati, unire categorie, aggiungere, aggregare, unire, dividere e trasporre i file ed eseguire molte altre operazioni.

Distribuzione elettronica. Con un semplice clic del mouse è possibile inviare report tramite posta elettronica oppure esportare tabelle e grafici in formato HTML per distribuirli in Internet o in una rete Intranet.

Aiuto in linea. Le esercitazioni dettagliate forniscono una panoramica completa del programma; l'Aiuto sensibile al contesto disponibile in tutte le finestre di dialogo consente di visualizzare informazioni relative a operazioni specifiche; le definizioni popup dei risultati delle tabelle pivot aiutano a comprendere i termini statistici; le Analisi guidate consentono di trovare rapidamente le procedure necessarie e gli Studi di casi offrono esempi pratici sull'utilizzo delle procedure statistiche e costituiscono un valido supporto all'interpretazione dei risultati.

Linguaggio a comandi. Benché la maggior parte delle attività possa essere eseguita con un semplice clic del mouse, in SPSS è disponibile anche un linguaggio a comandi che consente di salvare e automatizzare molte delle operazioni più comuni. Tale linguaggio offre inoltre alcune funzionalità a cui non è possibile accedere tramite i menu e le finestre di dialogo.

La documentazione completa sulla sintassi dei comandi è integrata nella Guida in linea ed è disponibile su un documento PDF a parte, *SPSS Command Syntax Reference*, anch'esso richiamabile dal menu ?.

Novità di SPSS 14.0

Gestione dei dati

- La possibilità di aprire contemporaneamente più sorgenti dati semplifica il confronto dei file di dati, consente di copiare dati e attributi da più file nonché di unire più sorgenti dati senza doverle salvare singolarmente come file di dati SPSS ordinato.
- Leggere e scrivere file di dati in formato Stata. La nuova versione consente di leggere file di dati Stata 4–8 e di scrivere file di dati Stata per le versioni 5–8. Per ulteriori informazioni, digitare Stata nella scheda Indice della Guida.
- Leggere dati da sorgenti dati SPSS Dimensions, compresi Quanvert, Quancept e mrInterview. Per ulteriori informazioni, vedere "Lettura di dati Dimensions" in Capitolo 3 a pag. 62.

- Leggere dati dalle sorgenti dati OLE DB. Per ulteriori informazioni, vedere "Selezione di una sorgente dati" in Capitolo 3 a pag. 33.
- Definire etichette descrittive per i valori di lunghezza massima pari a 120 byte (il limite precedente era 60 byte).
- Creare valori dati dalle etichette dei valori o usarli nella logica di conversione con la funzione `VALUELABEL`.
- Trovare e sostituire valori stringa con la funzione `REPLACE`.
- Specificare attributi per variabili e file di dati personalizzati con i comandi `VARIABLE ATTRIBUTE` che `DATAFILE ATTRIBUTE`.
- Scrivere dati nelle tabelle dei database e in altri formati utilizzando i nomi dei campi/delle colonne senza dover rispettare le regole di denominazione delle variabili di SPSS. L'opzione `SAVE TRANSLATE` è stata potenziata per permettere agli utenti di usare valori tra virgolette per i nomi di campi/colonne che contengono spazi, virgole e altri caratteri non utilizzabili per i nomi di variabili SPSS.
- Usare il nuovo sottocomando `SQL` del comando `SAVE TRANSLATE` per aggiungere nuove colonne alle tabelle dei database, modificare gli attributi delle colonnen delle tabelle dei database, unire tabelle ed eseguire altre operazionni con istruzioni SQL valide.

Grafici

- Usare la nuova interfaccia Generatore di grafici (menu Grafici) per creare grafici da modelli di grafici predefiniti o da parti di grafici (ad esempio sistemi di coordinate e barre).
- Creare tipi di grafici personalizzati tramite la sintassi dei comandi `GGRAPH` e `GPL`.

Nuove funzioni statistiche

- Il nuovo Expert Modeler in Trend permette di identificare automaticamente e di stimare il modello migliore per una o più serie temporali, eliminando il problema di dover ricercare il modello più idoneo manualmente. Expert Modeler può essere richiamato tramite la finestra di dialogo Modeler serie temporali oppure utilizzando la sintassi di comando con il comando `TSMODEL`.
- La nuova opzione Convalida dati permette di acquisire rapidamente un'istantanea dei dati e di applicare le regole di convalida che identificano i valori dei dati non validi. È possibile creare regole che escludono i valori fuori intervallo, i

valori mancanti o vuoti. È possibile anche salvare le variabili che registrano violazioni alle singole regole nonché il numero totale di violazioni alle regole per caso. Viene fornito un set limitato di regole predefinite da copiare o modificare. L'opzione Convalida dati può essere selezionata anche tramite la finestra di dialogo Convalida dati nel menu Dati oppure tramite la sintassi di comando utilizzando il comando VALIDATEDATA.

La nuova procedura Rilevamento anomali del modulo SPSS Data Validation permette di individuare osservazioni anomale che potrebbero influire sul risultato dei modelli predittivi. Alcune di queste osservazioni anomale rappresentano di fatto casi univoci e non sono indicati per la previsione, mentre altre sono dovute ad errori di inserimento dati, ovvero a casi in cui i valori sono tecnicamente "corretti" che non vengono quindi rilevati dalla procedura Convalida dati. L'opzione Rilevamento anomalie può essere selezionata nella finestra di dialogo Identifica casi anomali del menu Dati o tramite la sintassi dei comandi (comando DETECTANOMALY).

- La nuova procedura Unfolding multidimensionale (PREFSCAL) nell'opzione Categorie tenta di trovare la struttura in un insieme di misure di prossimità tra gli oggetti riga e colonna. Questo processo viene effettuato assegnando le osservazioni a posizioni specifiche in uno spazio concettuale ridotto, in modo tale che le distanze tra i punti nello spazio corrispondano il più possibile alle dissimilarità specificate. In questo modo si ottiene una rappresentazione dei minimi quadrati degli oggetti all'interno dello spazio dimensionale ridotto, che nella maggior parte dei casi aiuta a comprendere meglio i dati. Questa procedura può essere attualmente eseguita tramite la sintassi di comando PREFSCAL.
- La nuova procedura Selezione predittore (SELECTPRED) in SPSS Server è in grado di analizzare un numero elevato di variabili predittore categoriali e continue. La procedura seleziona il sottoinsieme più piccolo o lo usa per la creazione di modelli predittivi che non consentono l'uso di molti predittori. Questa procedura può essere attualmente eseguita tramite la sintassi di comando SELECTPRED.
- La nuova procedura Naive Bayes (NAIVEBAYES) in SPSS Server genera un modello semplice e stabile per la selezione e la classificazione dei predittori. Questa procedura può essere attualmente eseguita tramite la sintassi di comando NAIVEBAYES.
- Le funzionalità migliorate dei test di significatività del modulo aggiuntivo SPSS Tables permettono ora di eseguire test di significatività su subtotali e insiemi a risposta multipla.

- Sono disponibilii opzioni più flessibili per la definizione di insiemi a risposta multipla per le dicotomie multiple.

Uscita

- L'output della tabella pivot è ora disponibile per Classifica casi (`RANK`), Sostituisci valori mancanti (`RMV`) e Crea serie temporali (`CREATE`) nel sistema di base nonché per tutte le procedure dell'analisi Conjoint; per l'analisi loglineare completa di selezione del modello (`HILOGLINEAR`) nell'opzione Modelli avanzati; per l'analis Probit (`PROBIT`), la Stima dei pesi (`WLS`) e per i Minimi quadrati a 2 stadi (`2SLS`) nell'opzione Modelli di regressione.

Miglioramento delle prestazioni

- Le strutture delle tabelle che in passato richiedevano molto tempo per la creazione o che occupavano una grande quantità di memoria quando si utilizzava l'opzione Tabelle personalizzate (`CTABLES`) possono ora essere create in modo rapido ed efficiente.

Layout migliorato

- Le nuove icone della variabili forniscono informazioni più immediate sulle variabili compreso il livello di misurazione (nominale, ordinale, scala) e sul tipo di dati (stringa, numero, data, ora).
- Viene fornito un supporto completo per le impostazioni del desktop e dei temi di Windows XP.

Compatibilità di SPSS 14.0 con le versioni precedenti

Funzioni ANY e RANGE

Nelle versioni precedenti le funzioni `ANY` e `RANGE` restituivano un valore mancante solo se il primo argomento restituiva un valore mancante. Per uniformità con altre funzioni e altri calcoli, queste funzioni restituiscono ora un valore mancante se uno degli argomenti restanti è un valore mancante di sistema o definito dal'utente e se il valore del primo argomento non corrisponde agli altri argomenti non mancanti. Quindi:

```
COMPUTE newvar=ANY(var1, var2, var3)
```

è ora funzionamente equivalente a:

```
COMPUTE newvar=(var1=var2 or var1=var3).
```

Regressione logistica

Nelle versioni di SPSS precedenti, l'ordine dei valori stringa ricodificati era dipendente dall'ordine dei valori nel file di dati; ad esempio quando si registrava una variabile dipendente, il primo valore stringa trovato veniva ricodificato in 0 mentre il secondo valore stringa trovato veniva ricodificato in 1. La procedura è ora in grado di ricodificare le variabili stringa in modo che l'ordine dei valori ricodificati corrisponda all'ordine alfanumerico dei valori stringa. Quindi è possibile che la procedura ricodifichi le variabili stringa in modo diverso dalle versioni precedenti. La Regressione logistica può essere ora selezionata dall'opzione Modelli di regressione.

Funzionalità Macro

Le migliorie apportate all'utilità delle macro possono causare errori nei lavori che in precedenza non generavano errori. Nello specifico, per le sintassi elaborate con regole interattive, se una chiamata macro si verifica alla fine del comando e non è presente alcun elemento che termini il comando (ad esempio un punto o una riga vuota), il comando successivo dopo l'espansione della macro verrà interpretato come una riga di continuazione anziché come nuovo comando; come nel seguente esempio:

```
DEFINE !macro1() var1 var2 var3
!ENDDEFINE.
FREQUENCIES VARIABLES = !macro1
DESCRIPTIVES VARIABLES = !macro1.
```

In modalità interattiva il comando `DESCRIPTIVES` verrà interpretato come continuazione del comando `FREQUENCIES`, quindi non verrà eseguito nessuno dei due comandi.

Finestre

In SPSS sono disponibili diversi tipi di finestre:

Editor dei dati. Nell'Editor dei dati viene visualizzato il contenuto del file di dati. Nell'Editor dei dati è possibile creare nuovi file di dati o modificare quelli esistenti. Se ci sono più file di dati aperti, ciascun file di dati viene visualizzato una finestra a parte dell'Editor di dati.

Viewer. Tutti i risultati statistici, le tabelle e i grafici vengono visualizzati nel Viewer. È possibile modificare l'output e salvarlo per un successivo utilizzo. Una finestra del Viewer viene aperta automaticamente la prima volta che viene eseguita una procedura che genera output.

Bozza. Nella Bozza, l'output può essere visualizzato come testo semplice anzichè come tabelle pivot interattive.

Editor di tabelle pivot. L'output visualizzato nelle tabelle pivot può essere modificato in molti modi nell'Editor di tabelle pivot. È possibile modificare il testo, scambiare i dati contenuti in righe e colonne, aggiungere colore, creare tabelle multidimensionali e visualizzare o nascondere i risultati in modo selettivo.

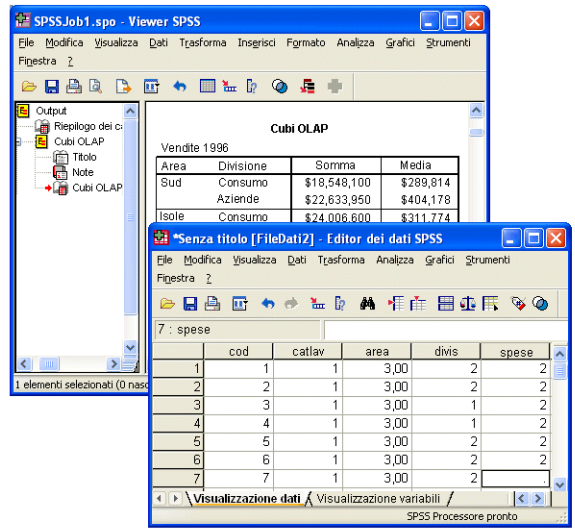
Editor dei grafici. È possibile modificare grafici ad alta risoluzione nelle finestre dei grafici. È possibile modificare i colori, selezionare diversi tipi o dimensioni di carattere, scambiare gli assi orizzontale e verticale, ruotare grafici a dispersione 3-D e perfino cambiare tipo di grafico.

Editor di output testuale. L'output testuale non visualizzato nelle tabelle pivot può essere modificato nell'Editor di output testuale. È possibile modificare l'output e modificare le caratteristiche del carattere (tipo, stile, colore, dimensioni).

Editor della sintassi. È possibile incollare le impostazioni della finestra di dialogo in una finestra della sintassi quando tali impostazioni vengono visualizzate sotto forma di sintassi. È possibile modificare la sintassi dei comandi per utilizzare particolari funzioni di SPSS non disponibili nelle finestre di dialogo. È possibile salvare tali comandi in un file da utilizzare nelle successive sessioni di SPSS.

Editor degli script. Le funzioni per la creazione di script e l'automazione OLE consentono di personalizzare e automatizzare numerose operazioni in SPSS. Utilizzare l'Editor degli script per creare e modificare script di base.

Figura 1-1
Editor dei dati e Viewer


SPSSJob1.spo - Viewer SPSS
File Modifica Visualizza Dati Trasforma Inserisci Formato Analizza Grafici Strumenti
Finestra ?
Output
Riepilogo dei c
Cubi OLAP
Titolo
Note
Cubi OLAP
Cubi OLAP
Vendite 1996
Area Divisione Somma Media
Sud Consumo $18,548,100 $289,814
Aziende $22,633,950 $404,178
Isole Consumo $24,006,600 $311,774
*Senza titolo [FileDati2] - Editor dei dati SPSS
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti
Finestra ?
7 : spese
cod catlav area divis spese
1 1 1 3,00 2 2
2 2 1 3,00 2 2
3 3 1 3,00 1 2
4 4 1 3,00 1 2
5 5 1 3,00 2 2
6 6 1 3,00 2 2
7 7 1 3,00 2 .
Visualizzazione dati
Visualizzazione variabili
SPSS Processore pronto

Confronto tra finestra designata e attiva

Se sono aperte più finestre del Viewer, l'output viene trasferito nella finestra **designata**. Se sono aperte più finestre dell'Editor della sintassi, la sintassi viene incollata nella finestra designata. Le finestre designate sono contrassegnate riportano

un simbolo più nell'icona della barra del titolo. È possibile cambiare le finestre designate in qualsiasi momento.

La finestra designata non deve essere confusa con la finestra **attiva**, che è la finestra correntemente selezionata. Se le finestre sono sovrapposte, la finestra attiva viene visualizzata in primo piano. Se si apre una nuova finestra, questa diventerà automaticamente la finestra attiva e la finestra designata.

Per cambiare la finestra designata del Viewer o dell'Editor della sintassi

- Fare clic in un punto qualsiasi della finestra per renderla attiva.
- Fare clic sul pulsante Designa finestra sulla barra degli strumenti (l'icona con il simbolo più).

  oppure

- Dai menu, scegliere:

  Strumenti
  Designa finestra

Nota: per le finestre dell'Editor dei dati, la finestra attiva determina il file di dati che verrà usato per i calcoli o le analisi successivi. In questo caso non è disponibile nessuna finestra "designata". Per ulteriori informazioni, vedere "Gestione di più sorgenti di dati" in Capitolo 6 a pag. 120.

Menu

La maggior parte delle operazioni che si possono eseguire in SPSS richiedono la selezione di comando da un menu. Ogni finestra di SPSS è dotata della propria barra dei menu in cui sono disponibili i comandi appropriati per quel tipo di finestra.

I menu Analizza e Grafici sono disponibili in tutte le finestre, quindi è possibile generare nuovo output senza necessità di cambiare finestra.

Barra di stato

Nella barra di stato, situata nella parte inferiore di ciascuna finestra di SPSS, vengono visualizzate le seguenti informazioni:

Stato del file dati dei comandi. Per ogni procedura o comando eseguito, verrà indicato il numero di casi già elaborati. Per le procedure statistiche che richiedono un'elaborazione iterativa, verrà visualizzato il numero di iterazioni.

Stato filtro. Se è stato selezionato un campione casuale o un sottoinsieme di casi per l'analisi, il messaggio Filtro attivo indica che è attivo un filtro e che non tutti i casi del file di dati verranno inclusi nell'analisi.

Stato della ponderazione. Il messaggio Peso attivo indica che per pesare i casi dell'analisi viene utilizzata una variabile di ponderazione.

Stato Distingui. Il messaggio Distingui attivo indica che il file di dati è stato suddiviso in gruppi distinti ai fini dell'analisi in base ai valori di una o più variabili di raggruppamento.

Per visualizzare o nascondere la barra di stato

- Dai menu, scegliere:
  Visualizza
  Barra di stato

Finestre di dialogo

La maggior parte dei comandi dei menu di SPSS consente di aprire finestre di dialogo, dove è possibile selezionare variabili e opzioni per l'analisi.

Le finestre di dialogo delle procedure statistiche e dei grafici hanno in genere due componenti di base:

Elenco di variabili sorgente. Elenco delle variabili del file di dati attivo. dove vengono visualizzati solo i tipi di variabile consentiti dalla procedura selezionata. L'utilizzo delle variabili di stringa corta e lunga è limitato in molte procedure.

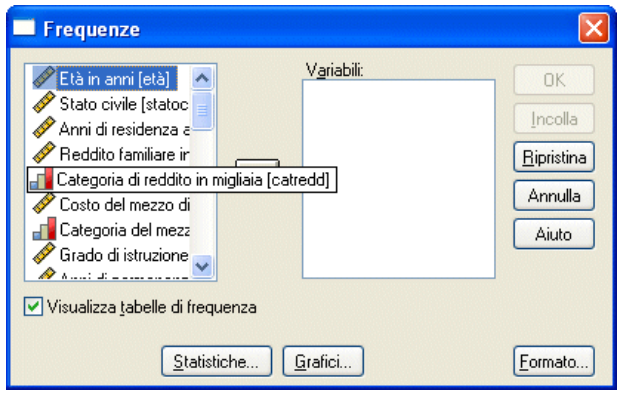
Elenco di variabili di destinazione. Le variabili scelte per l'analisi vengono visualizzate in una o più liste, ad esempio nelle liste delle variabili dipendenti e indipendenti.

Nomi ed etichette di variabili negli elenchi delle finestre di dialogo

È possibile visualizzare i nomi o le etichette descrittive delle variabili negli elenchi nelle finestre di dialogo.

- Per controllare la visualizzazione di nomi ed etichette di variabili, scegliere Opzioni dal menu Modifica di qualsiasi finestra.
- Per definire o modificare etichette di variabili, utilizzare Visualizzazione variabili nell'Editor dei dati.
- Per i dati importati da database, i nomi dei campi vengono utilizzati come etichette delle variabili.
- Nel caso di etichette lunghe, posizionare il puntatore del mouse sull'etichetta nell'elenco per visualizzarla per intero.
- Se le etichette delle variabili non sono definite, vengono visualizzati i nomi delle variabili.

Figura 1-2
Etichette di variabili visualizzate in una finestra di dialogo.



Controlli delle finestre di dialogo

Nella maggior parte delle finestre di dialogo di SPSS sono disponibili cinque controlli standard:

OK. Consente di eseguire la procedura. Dopo aver selezionato le variabili e impostato le eventuali opzioni aggiuntive, fare clic su OK per eseguire la procedura e chiudere la finestra di dialogo.

Incolla. Consente di generare la sintassi in base alle impostazioni della finestra di dialogo e di incollarla in una finestra della sintassi. È possibile personalizzare i comandi utilizzando funzioni aggiuntive di SPSS, non disponibili nelle finestre di dialogo.

Ripristina. Consente di deselezionare qualsiasi variabile nell'elenco delle variabili selezionate e di ripristinare tutte le impostazioni originali della finestra di dialogo e delle eventuali sottofinestre.

Annulla. Consente di annullare qualsiasi modifica apportata alle impostazioni della finestra di dialogo a partire dall'ultima apertura e quindi di chiudere la finestra stessa. Le impostazioni delle finestre di dialogo vengono mantenute per l'intera sessione attiva di SPSS. Le ultime impostazioni specificate nella finestra di dialogo verranno mantenute fino alla modifica successiva.

Aiuto. Visualizza l'aiuto contestuale. Questo controllo consente di accedere a una finestra dell'Aiuto in cui sono visualizzate informazioni relative alla finestra di dialogo corrente. Per richiedere aiuto nelle singole finestre di dialogo, fare clic con il pulsante destro del mouse sul controllo.

Sottofinestre di dialogo

Poiché la maggior parte delle procedure di SPSS è dotata di grande flessibilità, non sempre tutte le possibili opzioni possono essere incluse in un'unica finestra di dialogo. La finestra di dialogo principale in genere contiene il minimo di informazioni necessarie per eseguire una procedura. Nelle finestre di dialogo secondarie sono disponibili le opzioni aggiuntive.

I puntini (...) che seguono il nome di alcuni controlli disponibili nella finestra di dialogo principale indicano che facendo clic su tali comandi verranno visualizzate finestre di dialogo secondarie.

Selezione di variabili

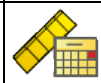
Per selezionare una singola variabile è sufficiente evidenziarla nell'elenco sorgente delle variabili e fare clic sul pulsante freccia rivolta a destra a lato della casella di riepilogo delle variabili di destinazione. Se esiste un solo elenco di destinazione delle variabili, è possibile fare clic sulle singole variabili per spostarle dall'elenco sorgente a quello di destinazione.

È inoltre possibile selezionare più variabili:

- Per selezionare più variabili raggruppate nell'elenco di variabili, fare clic sulla prima variabile e quindi, tenendo premuto Maiusc, fare clic sull'ultima variabile del gruppo.
- Per selezionare più variabili non raggruppate nell'elenco delle variabili, fare clic sulla prima variabile, quindi premere Ctrle e fare clic sulla variabile successiva e così via.

Icone dell'elenco Variabili

Le icone visualizzate accanto alla variabili negli elenchi delle finestre di dialogo forniscono informazioni sul tipo di variabile e sul livello di misurazione.

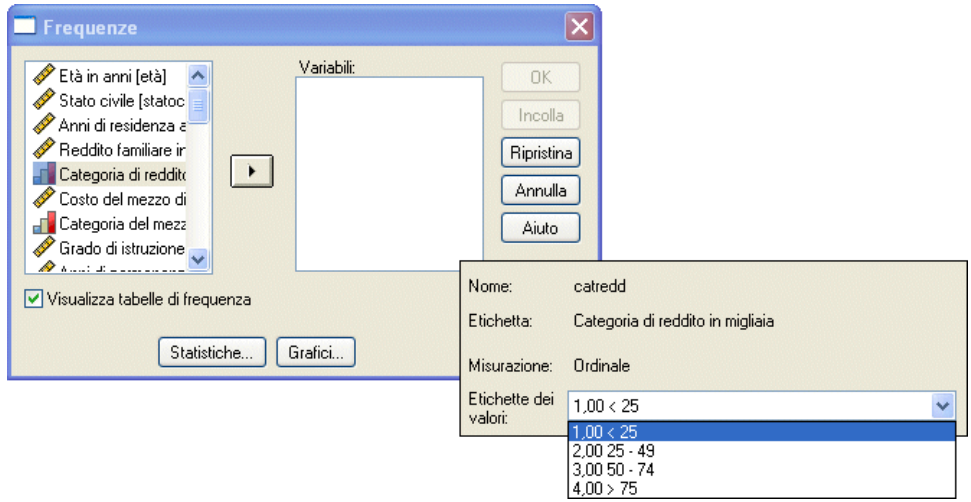
Livello di misurazione	**Data Type**			
	Numero	Stringa	Data	Ora
Scala		n/d		
Ordinale				
Nominale				

Per ottenere informazioni sulle variabili in una finestra di dialogo

- ▶ Fare clic con il pulsante destro del mouse su una variabile nell'elenco di variabili sorgente o di destinazione.

▶ Selezionare Informazioni variabili.

Figura 1-3
Informazioni sulle variabili



Passaggi fondamentali dell'analisi dei dati

In SPSS è facile analizzare i dati. È sufficiente:

Inserire i dati in SPSS. È possibile aprire un file di dati SPSS precedentemente salvato, leggere un foglio elettronico, un database o un file dati di testo oppure inserire i dati direttamente nell'Editor dei dati.

Selezionare una procedura. Selezionare una procedura dai menu per calcolare statistiche o creare un grafico.

Selezionare le variabili per l'analisi. Le variabili del file di dati verranno visualizzate in una finestra di dialogo relativa alla procedura.

Eseguire la procedura e verificare i risultati. I risultati vengono visualizzati nel Viewer.

Analisi guidate

Se non si ha familiarità con SPSS o con le procedure statistiche disponibili nel programma, è consigliabile consultare Analisi guidate, che offre un rapido metodo di apprendimento basato su domande semplici formulate in un linguaggio comune con esempi visivi che permettono di individuare le funzioni statistiche e per la creazione di grafici più adatte ai dati.

Per utilizzare Analisi guidate, dai menu di qualsiasi finestra di SPSS scegliere:

Aiuto
  Analisi guidate

In Analisi guidate vengono illustrate solo alcune procedure del modulo Base di SPSS. Viene fornita assistenza generica in relazione a molte delle tecniche statistiche di maggiore utilizzo.

Ulteriori informazioni su SPSS

Per informazioni generali sugli elementi di base di SPSS, vedere l'esercitazione in linea. Da qualsiasi barra dei menu di SPSS scegliere:

Aiuto
  Esercitazione

Capitolo
2

Aiuto

L'assistenza per l'uso del programma viene fornita in molti modi:

Menu ?. Il menu ? visualizzato nella maggior parte delle finestre di SPSS permette di accedere alla guida principale, alle esercitazioni e al materiale tecnico di riferimento.

- **Argomenti**. Consente di accedere alle schede Sommario, Indice e Cerca da utilizzare per trovare specifici argomenti dell'Aiuto.
- **Esercitazione.** Istruzioni guidate illustrate che spiegano come usare molte delle funzioni di base di SPSS. Non è necessario visualizzare tutta l'esercitazione dall'inizio alla fine. È possibile infatti scegliere gli argomenti che si desiderano visualizzare, selezionare gli argomenti desiderati in qualsiasi ordine ed utilizzare l'indice o il sommario per trovare argomenti specifici.
- **Studi di casi.** Esempi pratici che spiegano come creare vari tipi di analisi statistiche ed interpretarne i risultati. Sono inoltre disponibili i file di dati utilizzati negli esempi in modo da seguire questi ultimi e capire come sono stati ottenuti i risultati prodotti. È possibile scegliere le procedure specifiche sulle quali si desiderano maggiori informazioni dal sommario oppure ricercare gli argomenti desiderati nell'indice.
- **Analisi guidate.** Analisi di tipo guidato che consentono di trovare la procedura corretta. Appena si selezionano gli argomenti desiderati, le Analisi guidate aprono la finestra di dialogo corrispondente alle procedure per la creazione di statistiche, report o grafici che soddisfano i criteri selezionati. La funzione Analisi guidate consente di accedere alla maggior parte delle procedure statistiche e di reporting del modulo Base nonché a molte delle procedure per la creazione di grafici.

- **Sintassi.** Informazioni dettagliate sui riferimenti alla sintassi dei comandi sono disponibili in due formati: nella Guida integrata e su un documento PDF a parte in*SPSS Command Syntax Reference*, anch'esso consultabile tramite il menu ?.
- **Algoritmi statistici.** Gli algoritmi usati per la maggior parte delle procedure statistiche sono disponibili in un documento formato PDF consultabile tramite il menu ? e negli argomenti della Guida in linea relativi alla all'interfaccia della finestra di dialogo.

Aiuto contestuale. L'Aiuto contestuale può essere richiamato da più punti dell'interfaccia utente.

- **Pulsanti Aiuto nelle finestre di dialogo.** Nella maggior parte delle finestre di dialogo è disponibile un pulsante Aiuto che consente di passare direttamente a un argomento dell'Aiuto relativo a quella finestra di dialogo. Nell'argomento dell'Aiuto vengono fornite informazioni generali e collegamenti agli argomenti correlati.
- **Aiuto contestuale nelle finestre di dialogo.** Molte finestre di dialogo forniscono un aiuto contestuale su singoli comandi e funzioni. Fare clic con il pulsante destro del mouse su qualsiasi controllo di una finestra di dialogo e scegliere Guida rapida dal menu di scelta rapida per visualizzare una descrizione del controllo e delle sue modalità di utilizzo. Se l'opzione Guida rapida non è presente nel menu contestuale, è possibile che non sia disponibile per la finestra selezionata.
- **Aiuto contestuale nelle tabelle pivot.** Fare clic con il pulsante destro del mouse su un termine di una tabella pivot attiva nel Viewer e quindi scegliere Guida rapida dal menu di scelta rapida per visualizzare la relativa definizione.
- **Studi di casi.** Fare clic con il pulsante destro del mouse su una tabella pivot, quindi scegliere Studi di casi dal menu contestuale e passare direttamente all'esempio dettagliato della procedura che ha generato la tabella. Se l'opzione Studi di casi non è presente nel menu contestuale, è possibile che non sia disponibile per la procedura selezionata.
- **Sintassi dei comandi.** Nella finestra di una sintassi di comandi, spostare il cursore su qualsiasi punto del blocco della sintassi relativa ad un comando e premere F1 sulla tastiera. Viene visualizzato un grafico completo della sintassi dei comandi per il comando selezionato. La documentazione completa sulla sintassi dei

comandi può essere visualizzata selezionando i collegamenti nell'elenco di argomenti correlati o la scheda Contenuti dell'Aiuto.

- **Automazione OLE e creazione di script.** Nella finestra dello script (menu File, Nuovo o Apri, Script), è possibile selezionare ? per visualizzare ulteriori informazioni sul linguaggio degli script nonché sugli oggetti, i metodi e le proprietà di automazione OLE. Per richiamare la guida sensibile al contesto nella finestra di uno script, è sufficiente selezionare F1 o F2 (browser oggetto).

Impostazioni di Microsoft Internet Explorer

La maggior parte delle funzioni di aiuto di questa applicazione utilizzano una tecnologia basata su Microsoft Internet Explorer. Alcune versioni di Internet Explorer (compresa quella fornita con Microsoft XP, Service Pack 2) bloccano per impostazione predefinita ciò che considerano "contenuto attivo" nelle finestre di Internet Explorer del computer in uso. Questa impostazione predefinita può impedire di visualizzare alcune funzioni della Guida. Per visualizzare tutto il contenuto della Guida, è possibile modificare le impostazioni predefinite di Internet Explorer.

▶ Nel menu di Internet Explorer, selezionare:
Strumenti
Opzioni Internet...

▶ Fare clic sulla scheda Avanzate.

▶ Individuare la sezione Protezione.

▶ Selezionare la casella di controllo Consenti l'esecuzione di contenuto attivo sul computer.

Altre risorse

Se le informazioni desiderate non sono disponibili nella Guida, è possibile consultare anche altre risorse.

- **SPSS for Windows Developer's Guide.** Fornisce informazioni ed esempi relativi agli strumenti per sviluppatori di SPSS for Windows, compresi gli oggetti di automazione OLE, API di terze parti, DLL I/O, utilità Production e utilità per

gli script. Il documento *Developer's Guide* è disponibile in formato PDF nella directory *SPSS\developer* sul CD di installazione.

- **Sito Web del Supporto tecnico.** Per trovare una risposta a molti problemi comuni, è possibile visitare il sito Web *http://support.spss.com*. Per accedere al sito Web del supporto tecnico, sono necessari un ID di accesso e una password. Per informazioni su come richiedere l'ID e la password, consultare l'URL sopra specificato.

Utilizzo del sommario dell'Aiuto

▶ Dai menu di qualsiasi finestra, scegliere:

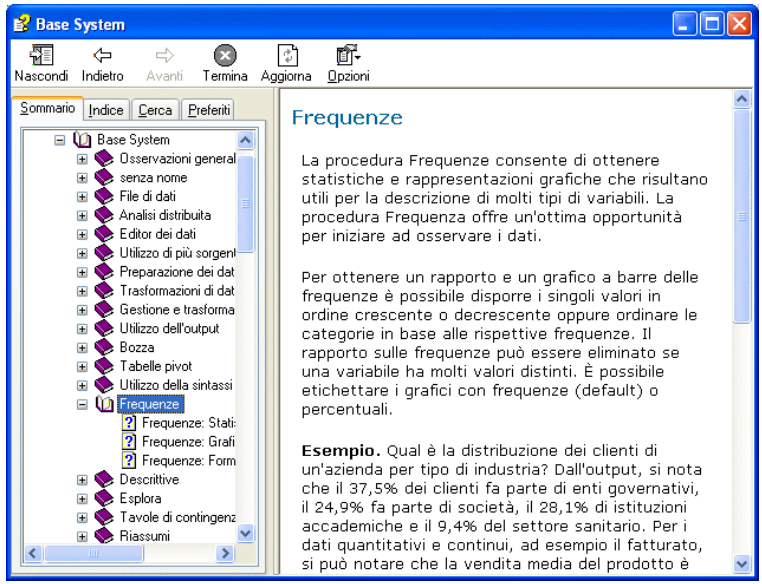
Aiuto
  Argomenti

▶ Fare clic sulla scheda Sommario.

▶ Fare doppio clic sulle voci contrassegnate da un'icona a forma di libro per espandere o comprimere il sommario.

▶ Fare clic sulla voce corrispondente all'argomento dell'Aiuto desiderato per visualizzarlo.

Figura 2-1
Finestra dell'Aiuto con visualizzazione della scheda Sommario


Base System
Nascondi
Indietro
Avanti
Termina
Aggiorna
Opzioni
Sommario
Indice
Cerca
Preferiti
Base System
Osservazioni general
senza nome
File di dati
Analisi distribuita
Editor dei dati
Utilizzo di più sorgent
Preparazione dei dat
Trasformazioni di dat
Gestione e trasforma
Utilizzo dell'output
Bozza
Tabelle pivot
Utilizzo della sintassi
Frequenze
Frequenze: Stati
Frequenze: Grafi
Frequenze: Form
Descrittive
Esplora
Tavole di contingenz
Riassumi
Frequenze
La procedura Frequenze consente di ottenere statistiche e rappresentazioni grafiche che risultano utili per la descrizione di molti tipi di variabili. La procedura Frequenza offre un'ottima opportunità per iniziare ad osservare i dati.
Per ottenere un rapporto e un grafico a barre delle frequenze è possibile disporre i singoli valori in ordine crescente o decrescente oppure ordinare le categorie in base alle rispettive frequenze. Il rapporto sulle frequenze può essere eliminato se una variabile ha molti valori distinti. È possibile etichettare i grafici con frequenze (default) o percentuali.
Esempio. Qual è la distribuzione dei clienti di un'azienda per tipo di industria? Dall'output, si nota che il 37,5% dei clienti fa parte di enti governativi, il 24,9% fa parte di società, il 28,1% di istituzioni accademiche e il 9,4% del settore sanitario. Per i dati quantitativi e continui, ad esempio il fatturato, si può notare che la vendita media del prodotto è

Utilizzo dell'Indice dell'Aiuto

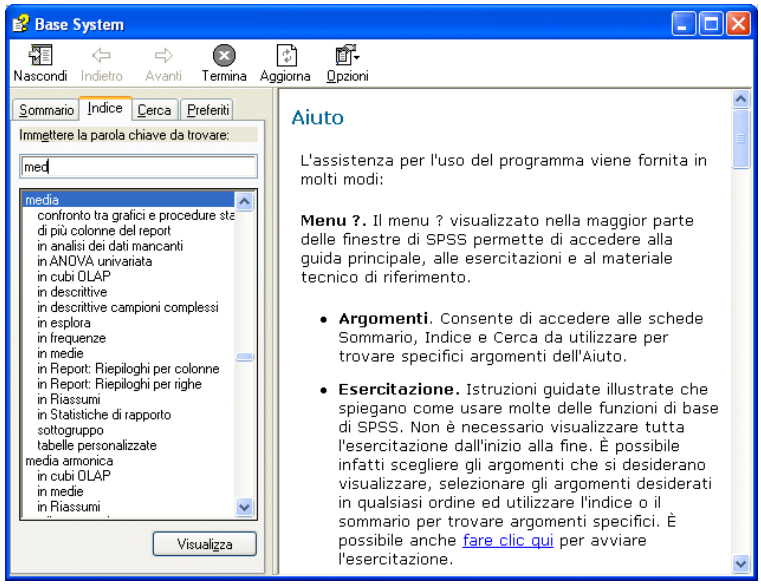
- Dai menu di qualsiasi finestra, scegliere:

  Aiuto
  Argomenti

- Fare clic sulla scheda Indice.
- Specificare il termine da ricercare nell'indice.
- Fare doppio clic sull'argomento desiderato.

  L'indice dell'Aiuto si avvale di un metodo di ricerca incrementale per trovare il testo specificato; verrà selezionata la voce dell'indice più vicina al termine indicato.

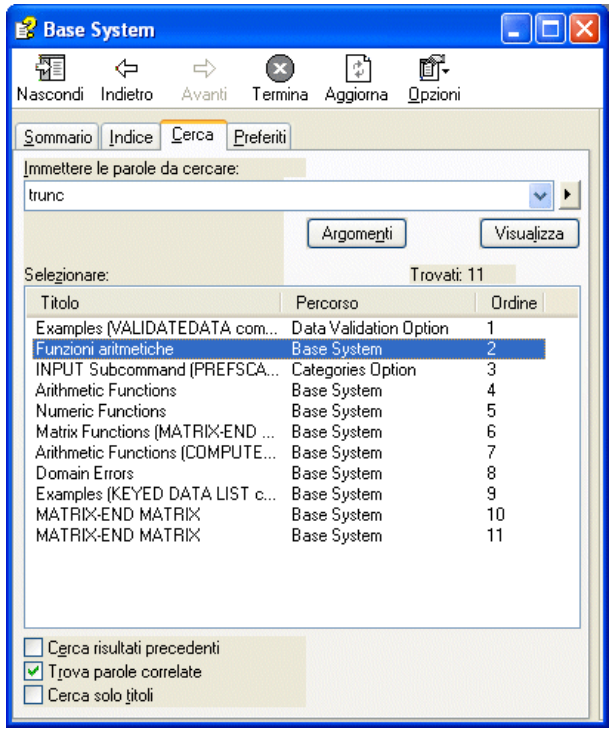
Figura 2-2
Scheda Indice e ricerca incrementale


Base System
Nascondi
Indietro
Avanti
Termina
Aggiorna
Opzioni
Sommario
Indice
Cerca
Preferiti
Immettere la parola chiave da trovare:
med
media
confronto tra grafici e procedure ste
di più colonne del report
in analisi dei dati mancanti
in ANOVA univariata
in cubi OLAP
in descrittive
in descrittive campioni complessi
in esplora
in frequenze
in medie
in Report: Riepiloghi per colonne
in Report: Riepiloghi per righe
in Riassumi
in Statistiche di rapporto
sottogruppo
tabelle personalizzate
media armonica
in cubi OLAP
in medie
in Riassumi
Visualizza
Aiuto
L'assistenza per l'uso del programma viene fornita in molti modi:
Menu ?. Il menu ? visualizzato nella maggior parte delle finestre di SPSS permette di accedere alla guida principale, alle esercitazioni e al materiale tecnico di riferimento.
Argomenti. Consente di accedere alle schede Sommario, Indice e Cerca da utilizzare per trovare specifici argomenti dell'Aiuto.
Esercitazione. Istruzioni guidate illustrate che spiegano come usare molte delle funzioni di base di SPSS. Non è necessario visualizzare tutta l'esercitazione dall'inizio alla fine. È possibile infatti scegliere gli argomenti che si desiderano visualizzare, selezionare gli argomenti desiderati in qualsiasi ordine ed utilizzare l'indice o il sommario per trovare argomenti specifici. È possibile anche fare clic qui per avviare l'esercitazione.

Uso della funzione di ricerca della Guida

La scheda Cerca permette di ricercare stringhe di testo complete, compresi titoli di argomenti e testi contenuti nei singoli argomenti o nell'indice analitico. Gli argomenti sono ordinati in base al numero di volte in cui il termine ricercato appare nell'argomento e/o nel relativo indice.

Figura 2-3
Scheda Ricerca della Guida



- È generalmente preferibile effettuare una ricerca di testo completa se è possibile ottimizzare i criteri di ricerca in modo da limitare il numero di argomenti visualizzato. Ad esempio per ricercare una funzione che tronchi i valori numerici, è preferibile digitare “funzione di troncamento” anziché “troncare” per visualizzare un elenco più mirato dei possibili argomenti rilevanti.
- Gli argomenti con nomi di comandi tra parentesi indicano che l’argomento della Guida si riferisce a una sintassi di comando. Gli argomenti senza nome di comando tra parentesi sono generalmente argomenti della Guida che si riferiscono a finestre di dialogo e ad elementi dell’interfaccia grafica.

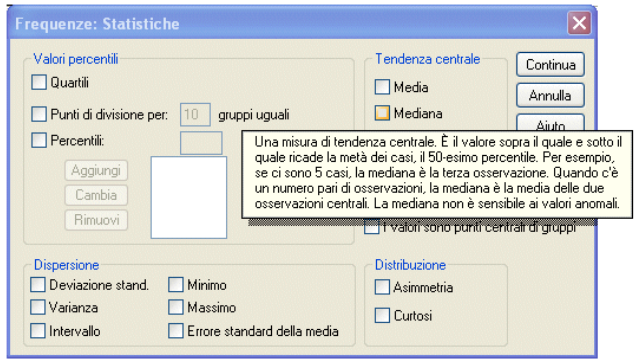
Visualizzazione di informazioni sui controlli delle finestre di dialogo

- Con il pulsante destro del mouse fare clic sul controllo della finestra di dialogo per il quale si desidera visualizzare informazioni.

- Scegliere Guida rapida dal menu di scelta rapida.

  In una finestra di messaggio verrà visualizzata una descrizione del controllo e delle sue modalità di utilizzo. Per visualizzare informazioni generali su una finestra di dialogo, utilizzare il pulsante Aiuto nella finestra stessa.

Figura 2-4
Visualizzazione dell'Aiuto per i controlli delle finestre di dialogo tramite il pulsante destro del mouse



Visualizzazione di informazioni sui termini delle tabelle pivot dell'output

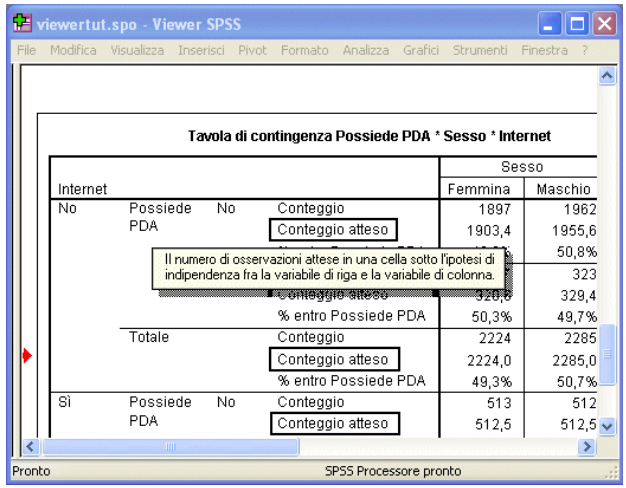
- Fare doppio clic nella tabella pivot per attivarla.

- Con il pulsante destro del mouse fare clic sul termine per il quale si desidera visualizzare una descrizione.

- Scegliere Guida rapida dal menu di scelta rapida.

  La definizione del termine verrà visualizzata in una finestra popup.

Figura 2-5
Visualizzazione dell'Aiuto per il glossario di una tabella pivot attivata tramite il pulsante destro del mouse



Utilizzo degli studi di casi

- Con il pulsante destro del mouse fare clic in una tabella pivot nella finestra del Viewer.
- Fare clic su Studi di casi nel menu di scelta rapida.

Per copiare il testo dell'Aiuto da una finestra popup

- Con il pulsante destro del mouse fare clic in un punto qualsiasi nella finestra popup.
- Scegliere Copia dal menu di scelta rapida.

  Verrà copiato l'intero testo della finestra popup.

Capitolo
3

File di dati

I file di dati sono disponibili in numerosi formati, molti dei quali possono essere gestiti in SPSS, ad esempio:

- Fogli elettronici creati con Excel e Lotus
- Le tabelle dei database di più sorgenti, compresi Oracle, SQLServer, Access, dBASE e altri
- File di testo tab delimitati e altri tipi di file di testo
- File di dati in formato SPSS creati in altri sistemi operativi.
- File di dati SYSTAT.
- File di dati SAS
- File di dati Stata

Apertura di un file di dati

Oltre ai file salvati in formato SPSS, è possibile aprire file di Excel, SAS, Stata e file delimitati da tabulazioni senza doverli convertire in un formato intermedio né inserire informazioni di definizione dei dati.

Per aprire file di dati

▶ Dai menu, scegliere:

File
  Apri
    Dati...

▶ Nella finestra di dialogo Apri file selezionare il file che si desidera aprire.

▶ Fare clic su Apri.

Se lo si desidera, è possibile:

- Leggere nomi di variabili dalla prima riga di fogli elettronici e file tab delimitati.
- Specificare un intervallo di celle da leggere per i fogli elettronici.
- Specificare il foglio del file di Excel da leggere (Excel 5 o versione successiva).

Tipi di file di dati

SPSS. Consente di aprire file di dati salvati in formato SPSS, inclusi SPSS per Windows, Macintosh, UNIX e in SPSS/PC+ per DOS.

SPSS/PC+. Consente di aprire file di dati SPSS/PC+.

SYSTAT. Consente di aprire file di dati SYSTAT.

Portabile SPSS. Consente di aprire file di dati in formato portabile SPSS. Il salvataggio di file in formato portabile richiede tempi notevolmente più lunghi del salvataggio in formato SPSS.

Excel. Consente di aprire i file di Excel.

Lotus 1-2-3. Consente di aprire i file salvati in formato 1-2-3 per le versioni 3.0, 2.0 o 1A di Lotus.

SYLK. Consente di aprire file di dati salvati in formato SYLK (Symbolic Link), utilizzato da molte applicazioni per fogli elettronici.

dBASE. Consente di aprire file in formato dBASE per dBASE IV, dBASE III, III PLUS o dBASE II. Ciascun caso rappresenta un record. Quando si salva un file in questo formato, le etichette delle variabili e dei valori e le specificazioni dei casi mancanti andranno perdute.

SAS, nome di file lungo. SAS versioni 7–9 per Windows, estensione lunga.

SAS, nome di file corto. SAS versioni 7–9 per Windows, estensione breve.

SAS v6 per Windows. SAS versione 6.08 per Windows e OS2.

SAS v6 per UNIX. SAS versione 6 per UNIX (Sun, HP, IBM).

Trasporto SAS. File di trasporto SAS.

Stata. Stata versioni 4–8.

Testo. File di testo ASCII.

Apertura file: opzioni

Leggi i nomi delle variabili. Per i fogli elettronici, è possibile leggere i nomi delle variabili dalla prima riga del file o dalla prima riga dell'intervallo specificato. I valori vengono convertiti in base alle necessità per creare nomi di variabili validi, inclusa la conversione degli spazi in caratteri di sottolineatura.

Foglio di lavoro. I file di Excel 5 o versioni successive possono contenere più fogli di lavoro. Per default, l'Editor dei dati legge il primo foglio di lavoro. Per leggere un altro foglio di lavoro, selezionare il foglio desiderato dall'elenco a discesa.

Intervallo. Per i file di dati di fogli elettronici è inoltre possibile leggere un intervallo di celle. Per specificare gli intervalli di celle è possibile utilizzare lo stesso metodo adottato nell'applicazione per fogli elettronici.

Lettura di file di Excel 5 o versione successiva

Per la lettura dei file di Excel 5 o versione successiva sono valide le seguenti regole:

Tipo e larghezza dei dati. Ciascuna colonna rappresenta una variabile. Il tipo e la larghezza dei dati di ciascuna variabile vengono determinati dal tipo e dalla larghezza dei dati del file di Excel. Se nella colonna sono presenti più tipi di dati (ad esempio, data e numerici), verrà impostato il tipo di dati stringa e tutti i valori verranno letti come valori stringa validi.

Celle vuote. Per le variabili numeriche, le celle vuote vengono convertite al valore mancante di sistema, indicato da un punto. Per le variabili stringa, uno spazio rappresenta un valore stringa valido. Le celle vuote, pertanto, vengono considerate come valori stringa validi.

Nomi delle variabili. Se la prima riga del file di Excel o la prima riga dell'intervallo specificato vengono lette come nomi delle variabili, i nomi non conformi alle regole di denominazione delle variabili verranno convertiti in nomi di variabili validi e i nomi originali verranno utilizzati come etichette delle variabili. Se i nomi delle variabili non vengono letti dal file di Excel, verranno assegnati i nomi di variabili predefiniti.

Per leggere file Excel di versioni precedenti e altri formati di fogli elettronici

Le regole riportate di seguito sono valide per la lettura di file di Excel antecedenti alla versione 5 e di file di dati di altri fogli elettronici:

Tipo e larghezza dei dati. Il tipo e la larghezza dei dati di ciascuna variabile vengono determinati dalla larghezza della colonna e dal tipo di dati della prima cella di dati della colonna. I valori di altri tipi verranno convertiti in valori mancanti di sistema. Se la prima cella di dati della colonna è vuota, viene utilizzato il tipo di dati globale di default per il foglio elettronico (in genere numerico).

Celle vuote. Per le variabili numeriche, le celle vuote vengono convertite al valore mancante di sistema, indicato da un punto. Per le variabili stringa, uno spazio rappresenta un valore stringa valido. Le celle vuote, pertanto, vengono considerate come valori stringa validi.

Nomi delle variabili. Se i nomi delle variabili non vengono letti dal foglio elettronico, le lettere delle colonne (*A*, *B*, *C*, ...) verranno utilizzate come nomi delle variabili nei file di Excel e di Lotus. Per i file SYLK e i file di Excel salvati in formato di visualizzazione R1C1, per i nomi delle variabili (*C1*, *C2*, *C3*, ...) verrà utilizzato il numero di colonna preceduto dalla lettera *C*.

Lettura di file dBASE

I file database sono molto simili ai file di dati in formato SPSS dal punto di vista logico. Per i file di dBASE sono valide le seguenti regole generali:

- I nomi dei campi vengono convertiti in nomi validi di variabili.
- I due punti utilizzati nei nomi di campo di dBASE verranno tradotti in caratteri di sottolineatura.
- Verranno inclusi i record contrassegnati per l'eliminazione che non sono stati effettivamente rimossi. Per i casi contrassegnati per l'eliminazione, verrà creata una nuova variabile stringa, *D_R*, contenente un asterisco.

Lettura di file Stata

Per i file di Stata sono valide le seguenti regole generali:

- **Nomi delle variabili.** I nomi delle variabili Strata vengono convertiti nei nomi delle variabili SPSS in maiuscolo/minuscolo. I nomi delle variabili Strata che risultano identici, ad eccezione del formato maiuscolo/minuscolo, vengono convertiti in nomi delle variabili SPSS validi aggiungendo un carattere di sottolineatura e una lettera sequenziale (*_A*, *_B*, *_C*, ..., *_Z*, *_AA*, *_AB*, ... e così via).
- **Etichette di variabile.** Le etichette di variabile Strata vengono convertite in etichette di variabile SPSS.
- **Etichette dei valori.** Le etichette dei valori Strata vengono convertiti in etichette dei valori SPSS, ad eccezione delle etichette dei valori Strata assegnate a valori mancanti "estesi".
- **Valori mancanti.** I valori mancanti "estesi" Strata vengono convertiti in valori mancanti di sistema.
- **Conversione di dati.** I valori Strata con formato data vengono convertiti in valori SPSS DATE con formato (g-m-a). I valori Strata con formato data "serie storica" (settimane, mesi, trimestri e così via) vengono convertiti nel formato numerico semplice (F), mantenendo il valore intero originale interno, ovvero il numero di settimane, mesi, trimestri e così via, a partire dall'inizio del 1960.

Lettura di file di database

Si possono leggere dati in qualsiasi formato di database per cui si dispone del driver appropriato. Nella modalità di analisi locale, è necessario che nel computer locale siano installati i driver richiesti. Nella modalità di analisi distribuita (disponibile con SPSS Server), i driver devono essere installati sul server remoto. Per ulteriori informazioni, vedere "Analisi distribuita" in Capitolo 4 a pag. 83.

Per leggere i file di database

▶ Dai menu, scegliere:

File
  Apri database
    Nuova query...

▶ Selezionare la sorgente dati.

- Se necessario (in base alla sorgente dati selezionata), specificare il file di database e/o di digitare un nome, una password e altre informazioni di login.
- Selezionare la tabella o le tabelle e i campi. (Per le sorgenti dati OLE DB, si può selezionare solo una tabella).
- Specificare le relazioni tra le tabelle.
- Oppure:
  - Specificare i criteri di selezione dei dati.
  - Aggiungere un prompt per l'input utente per creare una query di parametri.
  - Salvare la query prima di eseguirla.

Per modificare una query a database salvata

- Dai menu, scegliere:

  File
  Apri database
  Modifica query...
- Selezionare il file di query (**.spq*) da modificare.
- Seguire le istruzioni per la creazione di una nuova query.

Per leggere file di database con query salvate

- Dai menu, scegliere:

  File
  Apri database
  Esegui query...
- Selezionare il file di query (**.spq*) da eseguire.
- Se necessario (in base al file di database), immettere un nome e una password di login.
- Se la query ha un prompt incorporato, potrebbe risultare necessaria l'immissione di altre informazioni (ad esempio, il trimestre per cui si desidera ottenere le cifre relative alle vendite).

Selezione di una sorgente dati

Utilizzare questa prima schermata di Creazione guidata di query per selezionare il tipo di sorgente dati da leggere.

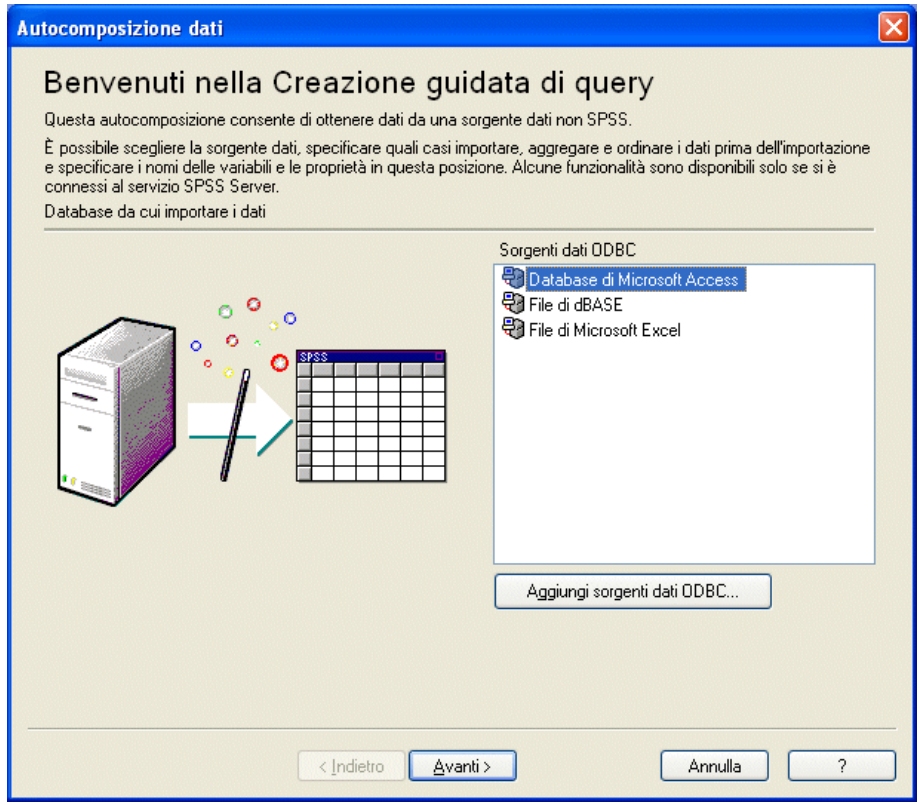
Sorgenti dati ODBC

Se non è stata configurata alcuna sorgente dati ODBC o se si desidera aggiungerne una nuova, fare clic su Aggiungi sorgente dati ODBC. Nella modalità di analisi distribuita (disponibile con SPSS Server) questo pulsante non è disponibile. Per aggiungere sorgenti dati nella modalità di analisi distribuita, rivolgersi all'amministratore di sistema.

Una sorgente dati ODBC include due informazioni fondamentali: il driver che verrà utilizzato per accedere ai dati e la posizione del database a cui si desidera accedere. Per specificare le sorgenti dati, è necessario che siano installati i driver appropriati e, nella modalità di analisi locale, è possibile installare i driver dal CD di SPSS:

- **SPSS Data Access Pack.** Installa i driver per un'ampia gamma di formati di database La funzione è disponibile nel menu di esecuzione automatica.
- **Microsoft Data Access Pack.** Installa i driver per i prodotti Microsoft, incluso Microsoft Access. Per installare Microsoft Data Access Pack, fare doppio clic su Microsoft Data Access Pack nella cartella Microsoft Data Access Pack del CD-ROM.

Figura 3-1
Finestra di dialogo Creazione guidata di query


Autocomposizione dati
Benvenuti nella Creazione guidata di query
Questa autocomposizione consente di ottenere dati da una sorgente dati non SPSS.
È possibile scegliere la sorgente dati, specificare quali casi importare, aggregare e ordinare i dati prima dell'importazione e specificare i nomi delle variabili e le proprietà in questa posizione. Alcune funzionalità sono disponibili solo se si è connessi al servizio SPSS Server.
Database da cui importare i dati
Sorgenti dati ODBC
Database di Microsoft Access
File di dBASE
File di Microsoft Excel
SPSS
Aggiungi sorgenti dati ODBC...
< Indietro
Avanti >
Annulla
?

Sorgenti dati OLE DB

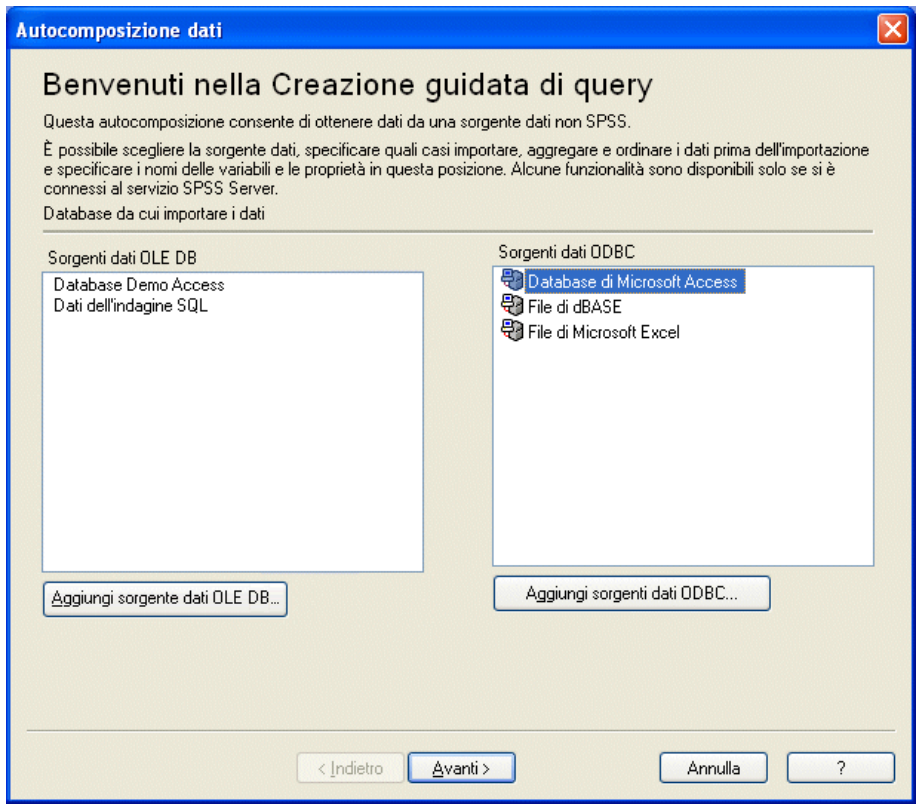
Per accedere alle sorgenti dati OLE DB, nel computer in cui si esegue SPSS devono essere installati gli elementi seguenti:

- .NET framework
- Dimension Data Model e OLE DB Access

È possibile installare le versioni di tali componenti compatibili con la presente versione di SPSS dal CD di installazione SPSS e sono disponibili nel menu di esecuzione automatica.

- È possibile aggiungere le sorgenti dati OLE DB solo in modalità di analisi locale. Per aggiungere delle sorgenti dati OLE DB in modalità di analisi distribuita in un server Windows, consultare l'amministratore di sistema.
- Nella modalità di analisi distribuita (disponibile con SPSS Server), le sorgenti dati OLE DB sono disponibili solo sui server Windows ed è necessario installare nel server .NET, Dimensions Data Model e OLE DB Access.

Figura 3-2
Creazione guidata di query con accesso alle sorgenti dati OLE DB


Autocomposizione dati
Benvenuti nella Creazione guidata di query
Questa autocomposizione consente di ottenere dati da una sorgente dati non SPSS.
È possibile scegliere la sorgente dati, specificare quali casi importare, aggregare e ordinare i dati prima dell'importazione e specificare i nomi delle variabili e le proprietà in questa posizione. Alcune funzionalità sono disponibili solo se si è connessi al servizio SPSS Server.
Database da cui importare i dati
Sorgenti dati OLE DB
Database Demo Access
Dati dell'indagine SQL
Aggiungi sorgente dati OLE DB...
Sorgenti dati ODBC
Database di Microsoft Access
File di dBASE
File di Microsoft Excel
Aggiungi sorgenti dati ODBC...
< Indietro
Avanti >
Annulla
?

Per aggiungere una sorgente dati OLE DB:

▶ Fare clic su Aggiungi sorgente dati OLE DB.

▶ Nella finestra di dialogo Proprietà di Data Link, fare clic sulla scheda Provider e selezionare il provider OLE DB.

- Fare clic su Avanti o sulla scheda Connessione.
- Selezionare il database immettendo la posizione della directory e il nome del database o facendo clic sul pulsante per selezionare un database. (Potrebbero essere necessari anche un nome utente e una password).
- Fare clic su OK dopo aver immesso le informazioni necessarie. (Per accertarsi che il database specificato sia disponibile, fare clic sul pulsante Prova connessione).
- Immettere un nome per le informazioni sulla connessione del database. (Tale nome verrà visualizzato nell'elenco delle sorgenti dati OLE DB disponibili).

Figura 3-3
Finestra di dialogo Salva con nome informazioni sulla connessione OLE DB

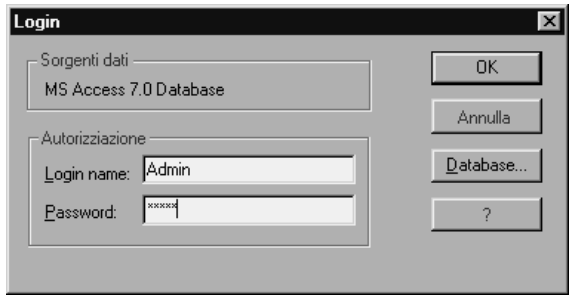
- Fare clic su OK.

L'operazione riconduce alla prima schermata di Creazione guidata di query, in cui è possibile selezionare il nome salvato dall'elenco di sorgenti dati OLE DB e proseguire con il passo successivo della procedura.

Accesso a un database

Se per aprire il database ODBC è necessaria una password, nella Creazione guidata di query verrà richiesto di specificare la password prima che sia possibile aprire la sorgente dati.

Figura 3-4
Finestra di dialogo Login

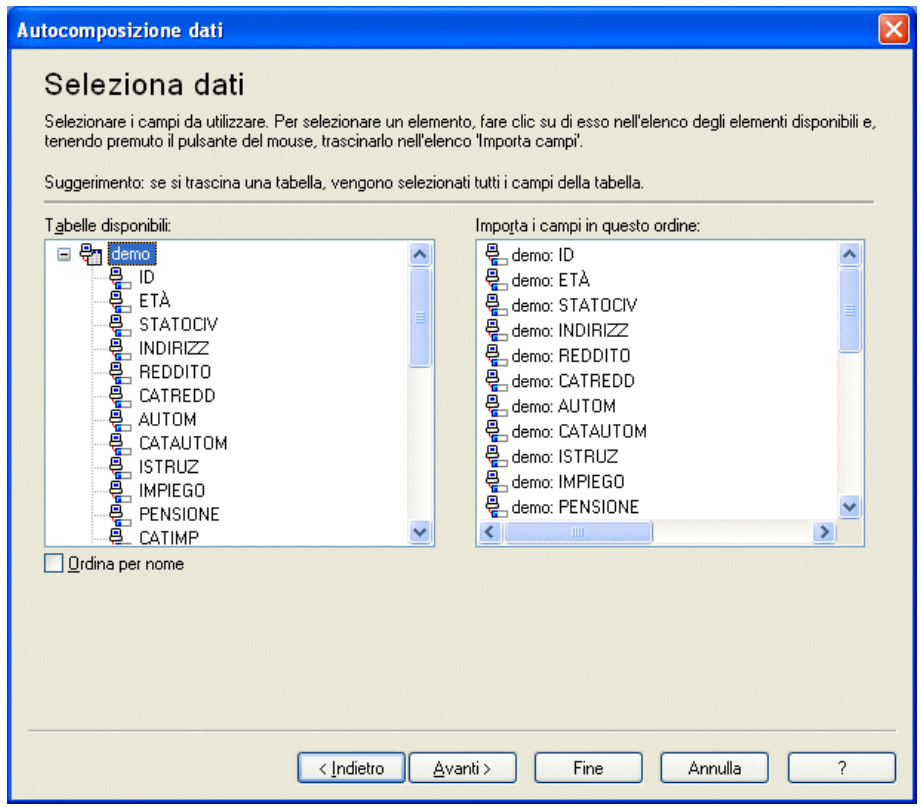


Selezione dei campi di dati

La finestra di dialogo Seleziona dei dati consente di controllare le tabelle e i campi che vengono letti. I campi del database (colonne) verranno letti come variabili.

Se in una tabella sono stati selezionati dei campi, tutti i campi saranno visibili nelle seguenti finestre della Creazione guidata di query, ma solo i campi selezionati in questa finestra di dialogo verranno importati come variabili. In questo modo è possibile creare join di tabelle e specificare i criteri utilizzando i campi che non vengono importati.

Figura 3-5
Finestra di dialogo Seleziona dati



Visualizzazione dei nomi di campo. Per elencare i campi di una tabella, fare clic sul segno più (+) a sinistra del nome della tabella. Per nascondere i campi, fare clic sul segno meno (–), posto a sinistra del nome della tabella.

Per aggiungere un campo. Fare doppio clic su un campo qualsiasi nell'elenco Tabelle disponibili oppure trascinarlo nell'elenco Importa i campi in questo ordine. Per riordinare i campi, è possibile trascinarli nell'elenco di campi.

Per rimuovere un campo. Fare doppio clic su un campo nell'elenco Importa i campi in questo ordine oppure trascinarlo nell'elenco Tabelle disponibili.

Ordina per nome. Se la casella di controllo è selezionata, nella Creazione guidata di query verranno visualizzati i campi disponibili in ordine alfabetico.

Nota: per le sorgenti dati OLE DB, si possono selezionare i campi solo da una una tabella. I join di tabelle multiple non sono supportati per le sorgenti dati OLE DB.

Creazione di una relazione tra tabelle

Nella finestra di dialogo Creazione di un collegamento è possibile definire le relazioni tra le tabelle relative alle sorgenti dati ODBC. Se vengono selezionati campi da più tabelle, è necessario definire almeno un join.

Figura 3-6
Finestra di dialogo Creazione di un collegamento

Creazione di relazioni. Per creare una relazione, trascinare un campo di una tabella nel campo a cui si desidera collegarlo. La Creazione guidata di query traccerà una **linea di collegamento** tra i due campi indicandone la relazione. I campi devono prevedere lo stesso tipo di dati.

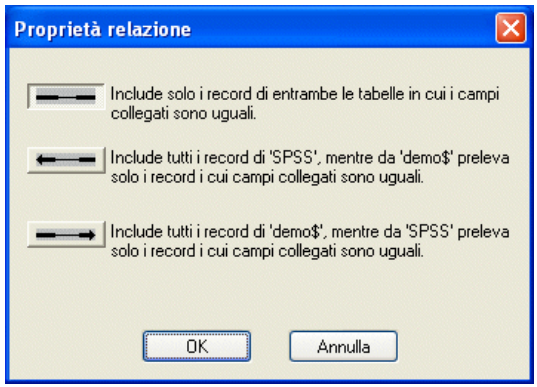
Collega tabelle. Questa procedura consente di eseguire automaticamente un tentativo di collegamento delle tabelle mediante le chiavi principali ed esterne oppure confrontando i nomi del campo con il tipo di dati.

Indicazione dei tipi di join. Se il driver utilizzato supporta i join esterni, è possibile specificare join interni, join esterni sinistri e join esterni destri. Per selezionare il tipo di join, fare doppio clic sulla linea di collegamento tra i campi. Verrà visualizzata la finestra di dialogo Proprietà relazione.

Proprietà relazione

In questa finestra di dialogo è possibile specificare il tipo di relazione tra due tabelle.

Figura 3-7
Finestra di dialogo Proprietà relazione


Proprietà relazione
Include solo i record di entrambe le tabelle in cui i campi collegati sono uguali.
Include tutti i record di 'SPSS', mentre da 'demo$' preleva solo i record i cui campi collegati sono uguali.
Include tutti i record di 'demo$', mentre da 'SPSS' preleva solo i record i cui campi collegati sono uguali.
OK
Annulla

Join interni. Un join interno include solo le righe in cui i campi correlati sono uguali. Nell'esempio, verranno incluse tutte le righe che riportano valori ID corrispondenti nelle due tabelle.

Join esterni. Oltre alla corrispondenza di uno a uno con i join interni, è possibile utilizzare i join esterni per unire le tabelle con uno scheda di corrispondenza di uno a molti. Per esempio, è possibile far corrispondere una tabella nella quale sono inclusi sono pochi record che rappresentano i valori dei dati e le etichette descrittive a essi

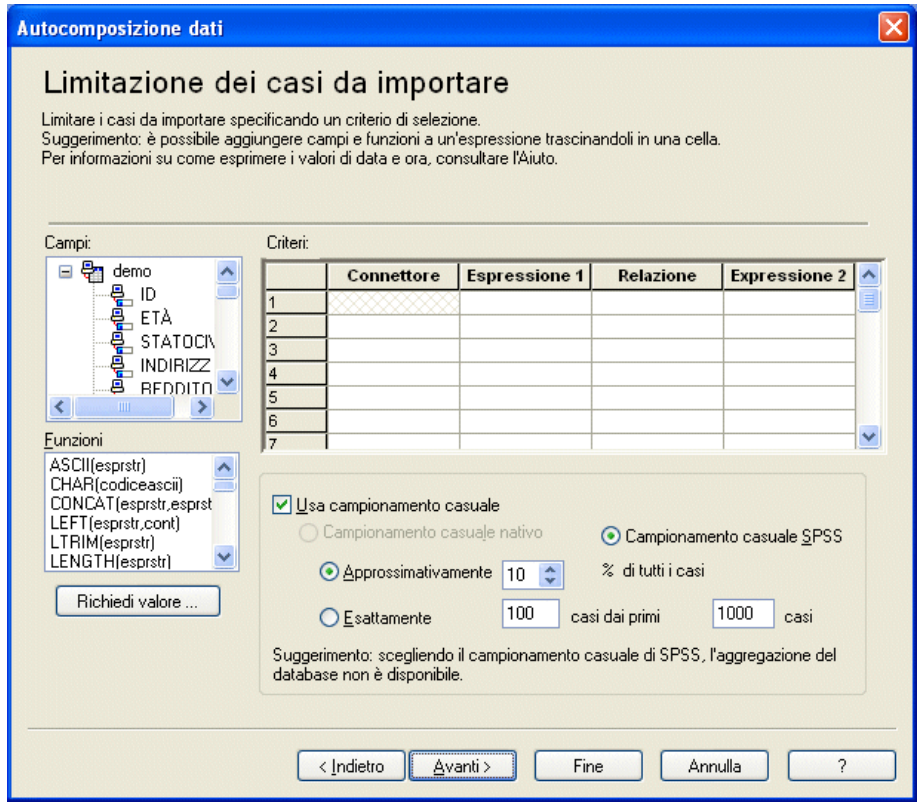
associate, con dei valori di una tabella che contiene centinaia o migliaia di record che rappresentano i rispondenti del sondaggio. Un join esterno sinistro include tutti i record della tabella a sinistra e solo i record della tabella a destra in cui i campi correlati sono uguali. In un join esterno destro, il join importa tutti i i record della tabella a destra e solo i record della tabella a sinistra in cui i campi correlati sono uguali.

Limitazione dei casi da importare

La finestra di dialogo Limitazione dei casi da importare consente di specificare i criteri per la selezione di sottoinsiemi di casi (righe). In genere, la limitazione dei casi consiste nel riempire la griglia dei criteri con dei criteri. I criteri consistono di due espressioni e di alcune relazioni reciproche. Le espressioni restituiscono un valore *vero*, *falso* o *mancante* per ciascun caso.

- Se il risultato è *vero*, il caso verrà selezionato.
- Se il risultato è *falso* o *mancante*, il caso non verrà selezionato.
- La maggior parte dei criteri utilizza uno o più dei sei operatori relazionali(<, >, <=, >=, = e <>).
- Le espressioni logiche possono includere nomi di campo, costanti, operatori aritmetici, funzioni numeriche e di altro tipo, nonché variabili logiche. È possibile utilizzare i campi che non si prevede di importare come variabili.

Figura 3-8
Finestra di dialogo Limitazione dei casi da importare



Per creare criteri personalizzati sono necessarie almeno due espressioni e una relazione che colleghi le espressioni.

▶ Per creare un'espressione, scegliere uno dei seguenti metodi:

- In una cella Espressione, immettere i nomi di campo, le costanti, gli operatori aritmetici e le funzioni numeriche e di altro tipo, nonché le variabili logiche.
- Fare doppio clic sul campo nell'elenco Campi.
- Trascinare il campo dall'elenco Campi in una cella Espressione.
- Selezionare un campo dal menu a discesa in qualsiasi cella Espressione attiva.

- Per selezionare un operatore relazionale (ad esempio = o >), posizionare il cursore sulla cella Relazione e digitare l'operatore o selezionarlo dal menu a discesa.

  Le date e le ore delle espressioni devono essere indicate in modo specifico (utilizzando anche le parentesi graffe indicate nell'esempio):

  - Le date in lettere devono essere specificate utilizzando il formato generico: {g aaaa-mm-gg}.
  - Le ore in lettere devono essere specificate utilizzando il formato generico: {t hh:mm:ss}.
  - Le date/ore in lettere (timestamp) devono essere specificate utilizzando il formato generico: {dt aaaa-mm-gg hh:mm:ss}.

  Funzioni. In SPSS sono disponibili diverse funzioni SQL di default di tipo aritmetico, logico, stringa, data e orario. È possibile trascinare una funzione dalla lista nell'espressione oppure inserire una funzione SQL valida. Per la lista delle funzioni SQL valide, consultare la documentazione del database. Per l'elenco delle funzioni standard disponibili, vedere:

  http://msdn.microsoft.com/library/en-us/odbc/htm/odbcscalar_functions.asp

  Usa campionamento casuale. Questa opzione consente di selezionare un campione casuale di casi dalla sorgente dati. Per le sorgenti dati di grandi dimensioni, è possibile limitare il numero dei casi a un campione piccolo e rappresentativo, riducendo in modo significativo i tempi di elaborazione. Il campionamento casuale nativo, se disponibile per la sorgente dati, risulta più veloce rispetto al campionamento casuale SPSS, perché quest'ultimo deve leggere l'intera sorgente dati per estrarre un campione casuale.

  - **Approssimativamente (Seleziona casi: campione casuale).** Estrae un campione casuale di numerosità approssimativamente uguale alla percentuale di casi specificata.
  - **Esattamente (Seleziona casi: campione casuale).** Seleziona esattamente il numero specificato di casi dai primi N casi specificati. Se il numero indicato è superiore al numero totale di casi nel file, il numero di casi estratti verrà ridotto proporzionalmente.

  Nota: Se si utilizza il campionamento casuale, l'aggregazione (disponibile in modalità distribuita in SPSS Server) non è disponibile.

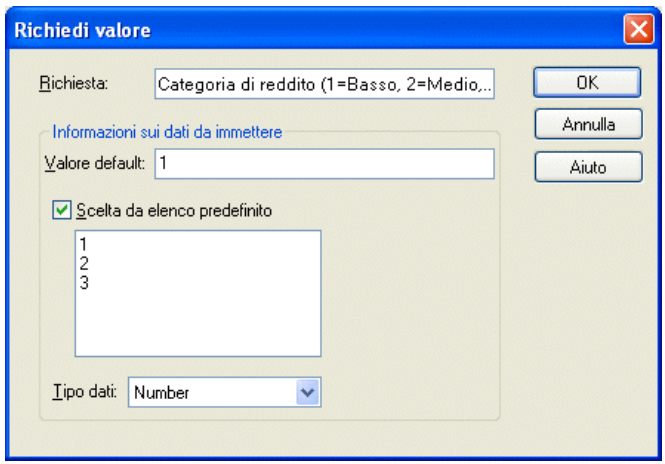
Richiedi valore. È possibile inserire nella query la richiesta di creare una **query di parametri**. Quando un utente esegue la query, verrà richiesto di immettere le informazioni (in base a quanto è stato specificato). Ciò potrebbe risultare utile per visualizzare gli stessi dati in modi diversi. È possibile, ad esempio, eseguire la stessa query per visualizzare le cifre di vendita relative a diversi trimestri fiscali.

- Posizionare il cursore su una cella Espressione e fare clic su Richiedi valore per creare una richiesta.

Creazione di una query di parametri

Utilizzare la finestra di dialogo Richiedi valore per creare una finestra di dialogo in cui vengono richieste informazioni all'utente ogni volta che si esegue una query. Questa procedura risulta utile per eseguire query per dati provenienti dalla stessa sorgente dati utilizzando criteri diversi.

Figura 3-9
Finestra di dialogo Richiedi valore



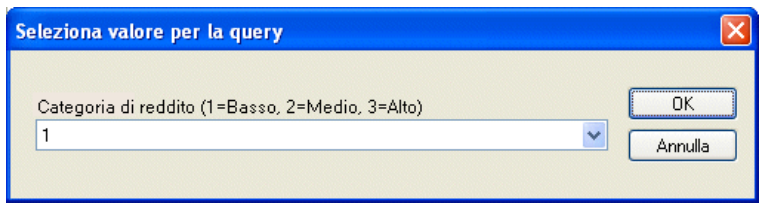
Per creare una richiesta di informazioni immettere una stringa di richiesta e un valore di default. Il testo della richiesta viene visualizzato ogni volta che un utente esegue la query. La stringa deve specificare il tipo di informazioni da immettere. Se l'utente non effettua la selezione da un elenco, la stringa deve fornire dei suggerimenti sulla formattazione dei dati. Esempio: Immettere un trimestre (t1, t2, t3, ...).

Scelta da elenco predefinito. Se la casella di controllo è selezionata, è possibile limitare i valori selezionabili dall'utente a quelli presenti in questo elenco. Accertarsi che i valori siano divisi da dei ritorni a capo.

Tipo dati. Selezionare il tipo di dati (numero, stringa o data).

Il risultato finale è simile al seguente:

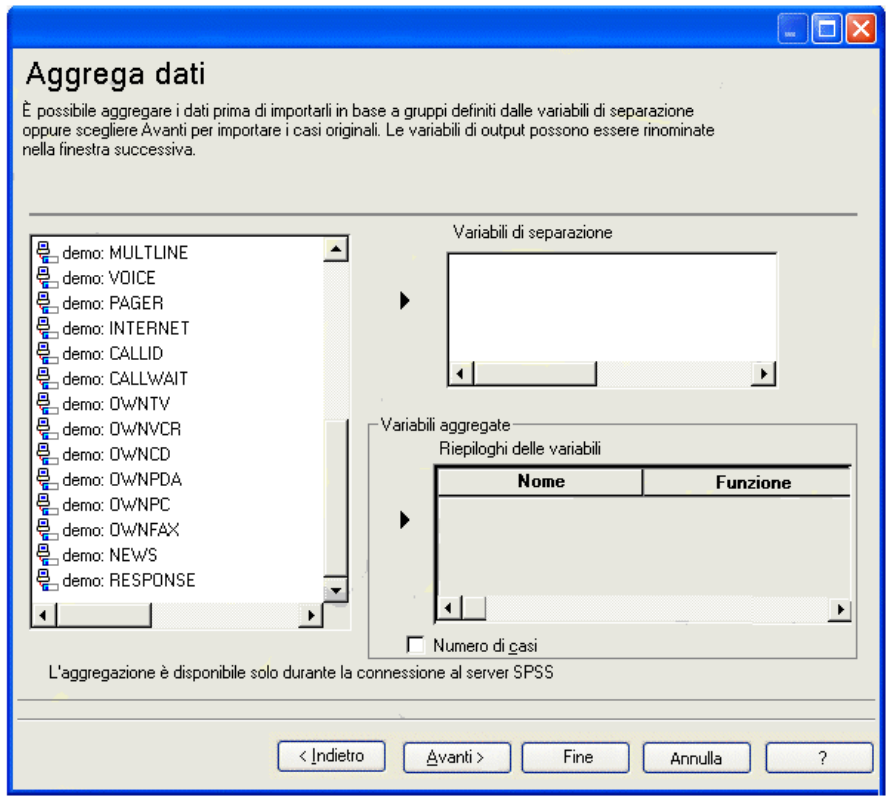
Figura 3-10
Finestra di dialogo del prompt definito dall'utente



Aggregazione dei dati

Se è attiva la modalità di analisi distribuita in connessione con un server remoto (disponibile con SPSS Server), è possibile aggregare i dati prima di eseguirne la lettura in SPSS.

Figura 3-11
Finestra di dialogo Aggrega dati



È inoltre possibile aggregare i dati dopo averne eseguito la lettura in SPSS; tuttavia, la pre-aggregazione può consentire di risparmiare tempo in caso di sorgenti dati di grandi dimensioni.

- Per creare dati aggregati, selezionare una o più variabili di separazione che definiscano le modalità di raggruppamento dei casi.
- Selezionare una o più variabili aggregate.
- Selezionare una funzione di aggregazione per ciascuna variabile aggregata.
- Se lo si desidera, è possibile creare una variabile che contenga il numero di casi di ciascun gruppo di interruzione.

Nota: se si usa il campionamento casuale, l'aggregazione non è disponibile.

Per definire una variabile

Nomi di variabili ed etichette. Come etichetta di variabile verrà utilizzato il nome completo del campo del database (colonna). A meno che non venga modificato il nome della variabile, la Creazione guidata di query assegna i nomi di variabile a ciascuna colonna del database in uno dei due seguenti modi:

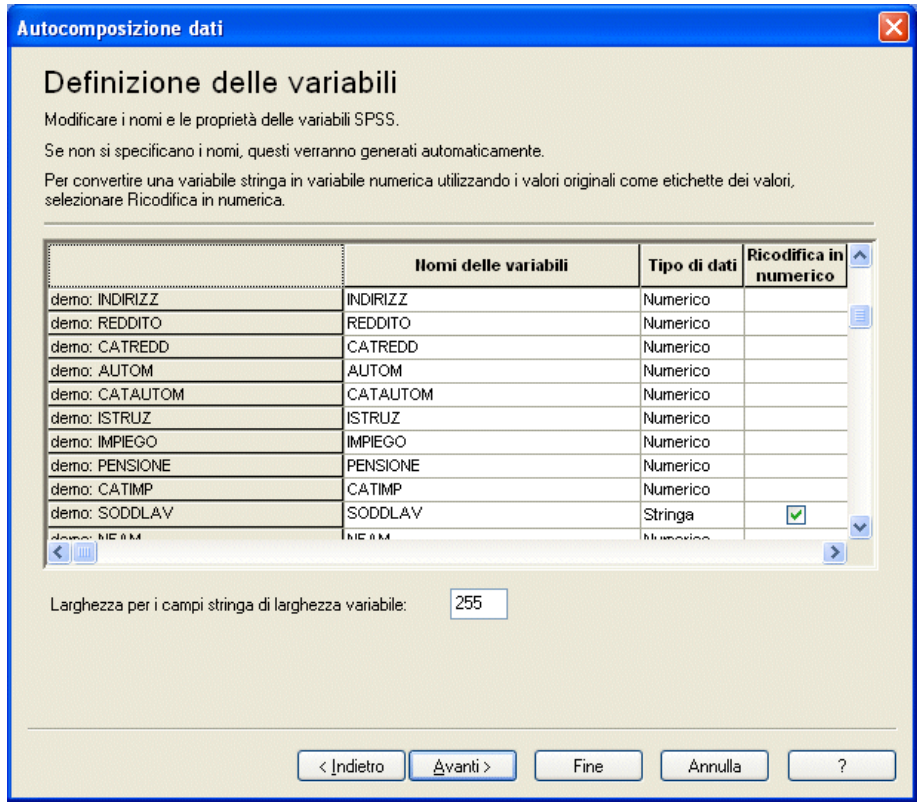
- Se il nome del campo del database forma un nome di variabile valido e univoco, questo verrà utilizzato come nome della variabile.
- Se il nome del campo del database non forma un nome di variabile valido e univoco, questo verrà creato automaticamente.

Fare clic su una cella per modificare il nome della variabile.

Conversione di stringhe in valori numerici. Selezionare la casella Ricodifica in numerica per la variabile stringa che si desidera convertire automaticamente in variabile numerica. I valori stringa verranno convertiti in valori interi consecutivi in base all'ordine alfabetico dei valori originali. I valori originali verranno mantenuti come etichette di valori per le nuove variabili.

Larghezza dei campi delle stringhe a larghezza variabile. L'opzione consente di controllare la larghezza dei valori delle stringhe a larghezza variabile. Per impostazione predefinita, la larghezza è 255 byte, sebbene vengano letti solo i primi 255 byte (generalmente i 255 caratteri nelle lingue a un byte). La larghezza massima è 32.767 byte. Sebbene sia preferibile non troncare i valori stringa, è consigliabile anche non specificare un valore alto se non è necessario perché questo renderà inefficiente SPSS.

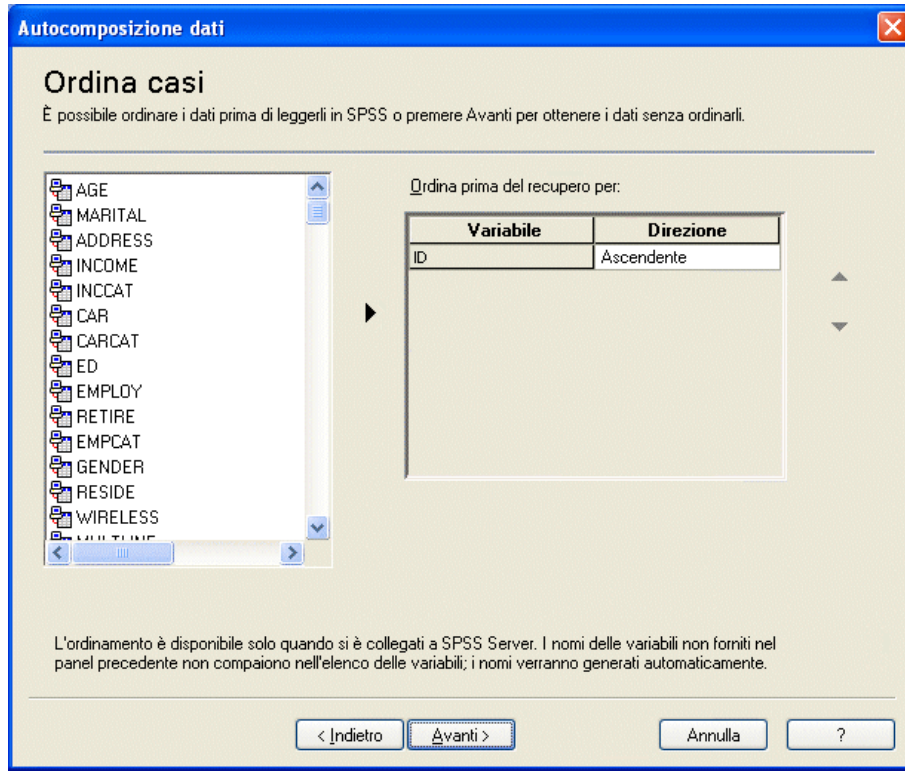
Figura 3-12
Finestra di dialogo Definizione delle variabili.



Ordinamento dei casi

Se è attiva la modalità di analisi distribuita in connessione con un server remoto (disponibile con SPSS Server), è possibile ordinare i dati prima di eseguirne la lettura in SPSS.

Figura 3-13
Finestra di dialogo Ordina casi.



È inoltre possibile ordinare i dati dopo averne eseguito la lettura in SPSS; tuttavia, il pre-ordinamento può consentire di risparmiare tempo in caso di sorgenti dati di grandi dimensioni.

Risultati

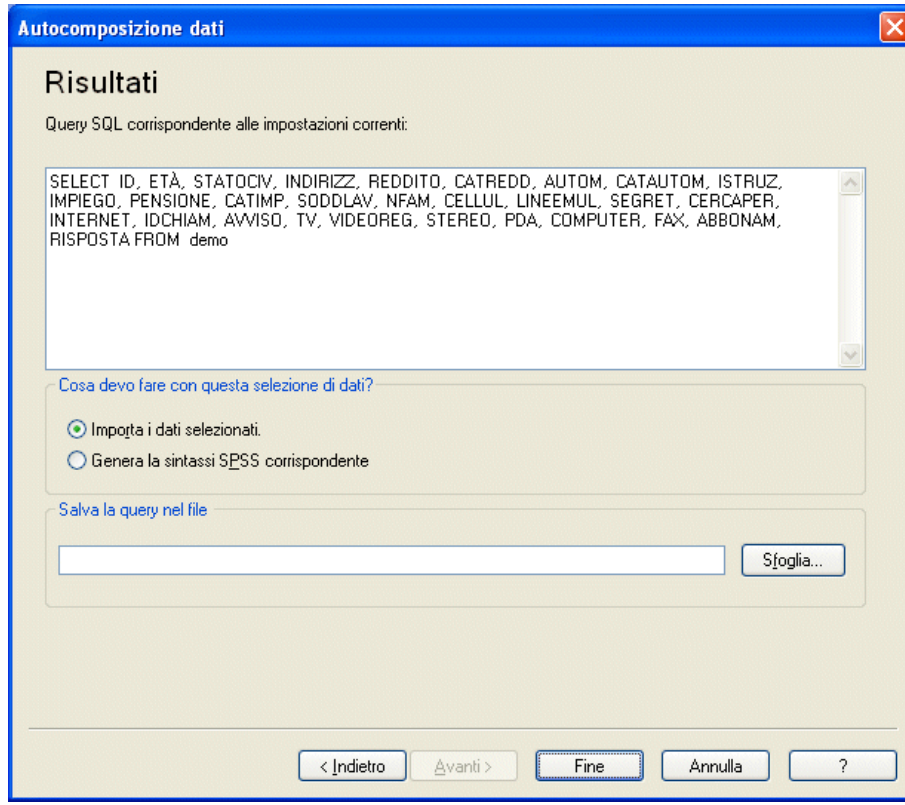
Nella finestra di dialogo Risultati viene visualizzata l'istruzione SQL Select della query.

- È possibile modificare l'istruzione SQL Select prima di eseguire la query, ma se si fa clic sul pulsante Indietro per apportare modifiche nei passi precedenti, le modifiche all'istruzione Select verranno perse.

- Per salvare la query e renderla disponibile per utilizzi futuri, utilizzare la sezione Salva la query nel file.
- Per incollare la sintassi completa di `GET DATA` in una finestra di sintassi, selezionare Genera la sintassi SPSS corrispondente. Se si copia e si incolla l'istruzione Select dalla finestra Risultati, la sintassi dei comandi richiesta non verrà incollata.

Nota: La sintassi incollata contiene uno spazio bianco prima della chiusura delle virgoletta su ciascuna riga dell'SQL generato dalla procedura guidata. Questi spazi bianchi non sono superflui. Al momento dell'elaborazione del comando, tutte le righe dell'istruzione SQL vengono unite letteralmente. Se non ci fosse lo spazio, non ci sarebbe alcuna divisione tra l'ultimo carattere della prima riga e il primo carattere di quella successiva.

Figura 3-14
Finestra di dialogo Risultati


Autocomposizione dati
Risultati
Query SQL corrispondente alle impostazioni correnti:
SELECT ID, ETÀ, STATOCIV, INDIRIZZ, REDDITO, CATREDD, AUTOM, CATAUTOM, ISTRUZ, IMPIEGO, PENSIONE, CATIMP, SODDLAV, NFAM, CELLUL, LINEEMUL, SEGRET, CERCAPER, INTERNET, IDCHIAM, AVVISO, TV, VIDEOREG, STEREO, PDA, COMPUTER, FAX, ABBONAM, RISPOSTA FROM demo
Cosa devo fare con questa selezione di dati?
Importa i dati selezionati.
Genera la sintassi SPSS corrispondente
Salva la query nel file
Sfoglia...
< Indietro
Avanti >
Fine
Annulla
?

Importazione guidata di testo

Importazione guidata di testo consente di leggere i file di testo formattati in vari modi:

- File tab delimitati
- File delimitati da spazi
- File delimitati da virgole
- File in formato di campo fisso

Per i file delimitati è inoltre possibile specificare altri caratteri come delimitatori da inserire tra i valori ed è possibile specificare più delimitatori.

Per leggere i file di dati di testo

- Dai menu, scegliere:

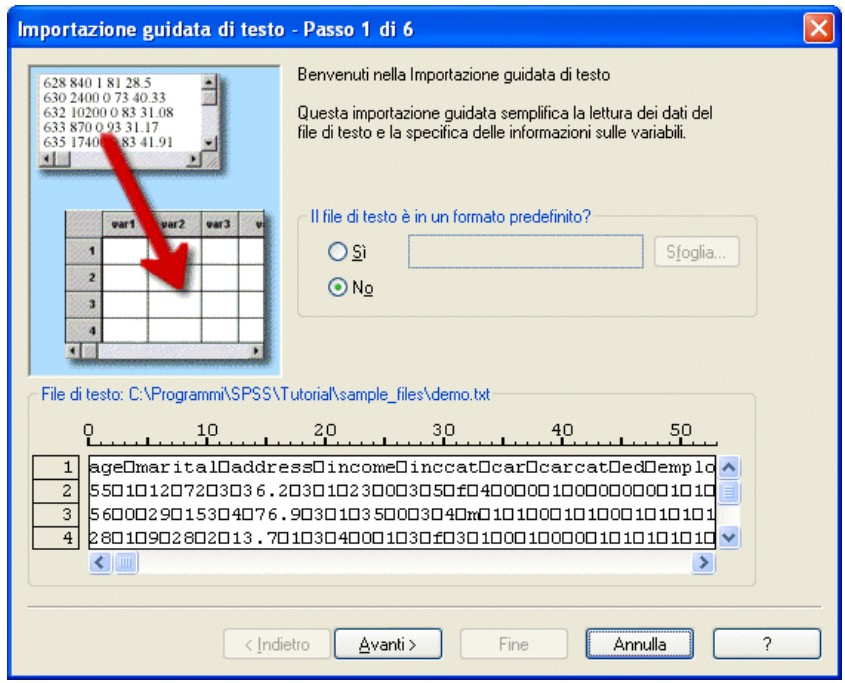
  File
  Leggi dati testuali...

- Selezionare il file di testo nella finestra di dialogo Apri file.

- Eseguire i passi di Importazione guidata di testo per definire il metodo di lettura dei file.

Importazione guidata di testo: Passo 1

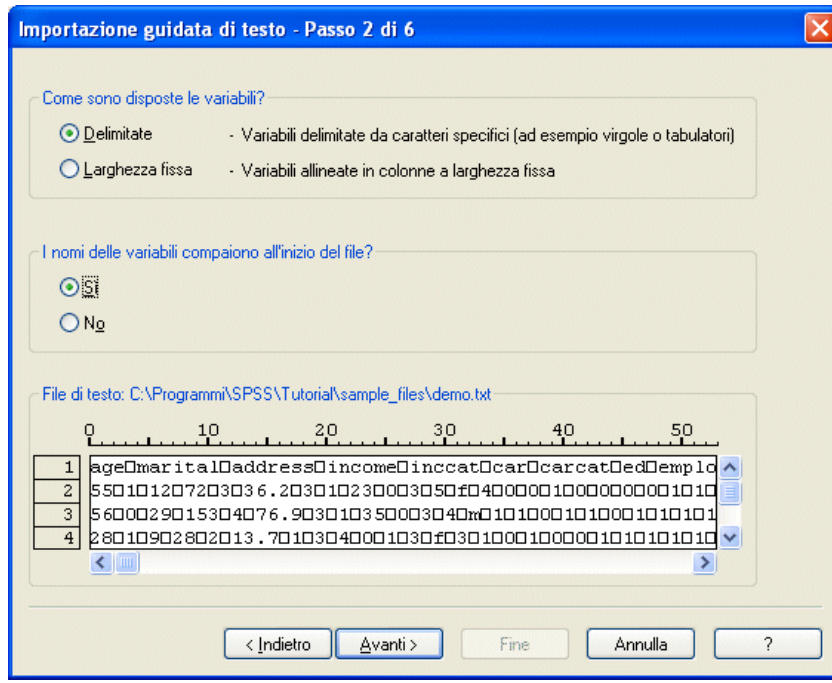
Figura 3-15
Importazione guidata di testo: Passo 1


Importazione guidata di testo - Passo 1 di 6
Benvenuti nella Importazione guidata di testo
Questa importazione guidata semplifica la lettura dei dati del file di testo e la specifica delle informazioni sulle variabili.
Il file di testo è in un formato predefinito?
Sì
No
Sfoglia...
File di testo: C:\Programmi\SPSS\Tutorial\sample_files\demo.txt
< Indietro
Avanti >
Fine
Annulla
?

Il file di testo viene visualizzato in una finestra di anteprima. È possibile applicare un formato predefinito (salvato precedentemente con Importazione guidata di testo) o eseguire i passi della procedura guidata per definire il metodo di lettura dei dati.

Importazione guidata di testo: Passo 2

Figura 3-16
Importazione guidata di testo: Passo 2


Importazione guidata di testo - Passo 2 di 6
Come sono disposte le variabili?
Delimitate - Variabili delimitate da caratteri specifici (ad esempio virgole o tabulatori)
Larghezza fissa - Variabili allineate in colonne a larghezza fissa
I nomi delle variabili compaiono all'inizio del file?
Sì
No
File di testo: C:\Programmi\SPSS\Tutorial\sample_files\demo.txt
< Indietro
Avanti >
Fine
Annulla
?

Nel passo 2 vengono fornite informazioni sulle variabili. Le variabili sono simili ai campi di un database. Ad esempio, ogni voce di un questionario è una variabile.

Come sono disposte le variabili? Per leggere correttamente i dati, Importazione guidata di testo deve poter determinare dove finiscono i valori dei dati di una variabile e iniziano i valori di quella successiva. La disposizione delle variabili determina il metodo utilizzato per separare una variabile dalla successiva.

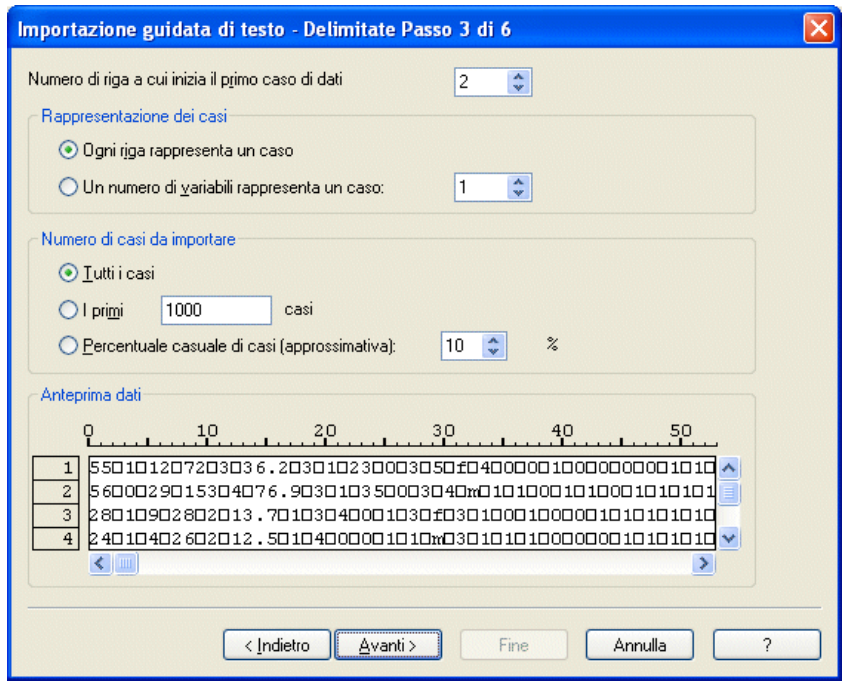
- **Delimitate.** Per separare le variabili possono essere usati spazi, virgole, tabulazioni o altri caratteri. Le variabili di ciascun caso vengono registrate nello stesso ordine, ma non sempre nelle stesse posizioni di colonna.
- **Larghezza fissa.** Ciascuna variabile viene registrata nella stessa posizione della colonna e nello stesso record (riga) per ciascun caso del file di dati. Non è richiesto nessun delimitatore tra le variabili. In molti file di dati di testo generati automaticamente potrebbe sembrare che i valori dei dati siano disposti senza

alcun simbolo di divisione. La posizione della colonna determina la variabile che verrà letta.

I nomi delle variabili compaiono all'inizio del file? Se la prima riga del file di dati contiene etichette descrittive delle variabili, è possibile utilizzare queste etichette come nomi di variabili. I valori non conformi alle regole di denominazione delle variabili vengono convertiti in nomi di variabile validi.

Importazione guidata di testo: Passo 3 (File delimitati)

Figura 3-17
Importazione guidata di testo: Passo 3 (per file delimitati)



Nel passo 3 vengono fornite informazioni sui casi. I casi sono simili ai record di un database. Ad esempio, ogni persona che risponde a un questionario rappresenta un caso.

Numero di riga a cui inizia il primo caso di dati. Indica la prima riga del file di dati che contiene valori. Se la prima riga o le prime righe del file di dati contengono etichette descrittive o testo di altro tipo che non rappresenta un valore, questa *non*sarà la riga 1.

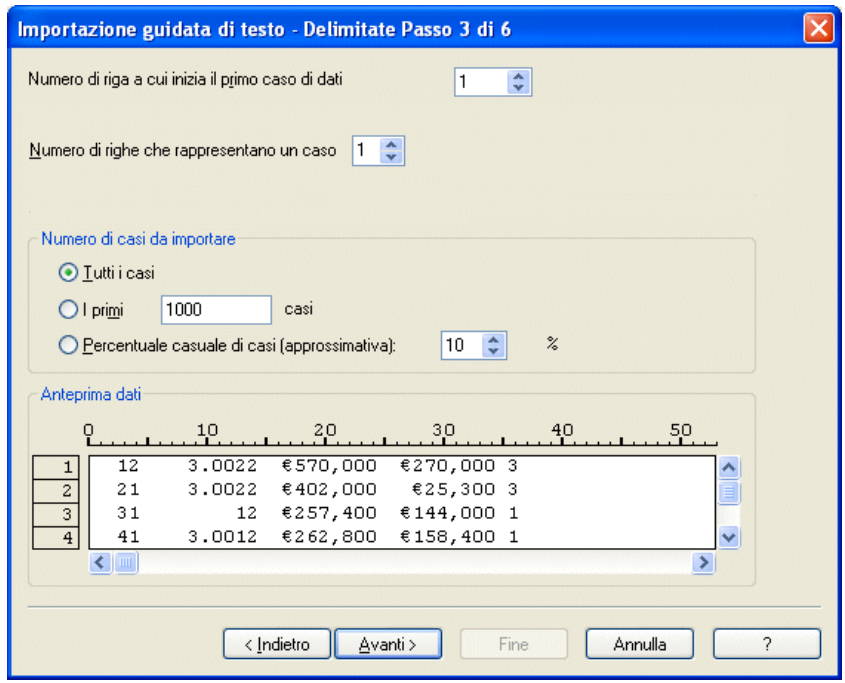
Rappresentazione dei casi. Controlla il modo in cui Importazione guidata di testo determina la fine di ogni caso e l'inizio di quello successivo.

- **Ogni riga rappresenta un caso.** Ogni riga contiene un solo caso. È abbastanza comune che ogni caso sia inserito su una sola riga, anche se questa potrebbe essere particolarmente lunga per i file di dati con un elevato numero di variabili. Se non tutte le righe contengono lo stesso numero di valori, il numero di variabili per ciascun caso è determinato dalla riga con il maggior numero di valori. Ai casi con meno valori sono assegnati valori mancanti per le variabili aggiuntive.
- **Un numero di variabili rappresenta un caso.** Il numero di variabili specificato per ogni caso indica a Importazione guidata di testo dove interrompere la lettura di un caso e iniziare quella del caso successivo. È possibile che più casi siano contenuti nella stessa riga e i casi possono anche iniziare a metà di una riga e proseguire nella riga successiva. Importazione guidata di testo determina la fine di ogni caso in base al numero di valori letti, indipendentemente dal numero di righe. Ogni caso deve contenere i valori dei dati o i valori mancanti indicati dai delimitatori per tutte le variabili, altrimenti i file di dati verranno letti in modo non corretto.

Numero di casi da importare. È possibile importare tutti i casi del file di dati, solo i primi *n* casi (*n* (n è un numero specificato dall'utente) o un campione casuale determinato in base a una percentuale specificata. Poiché viene eseguito un processo indipendente di decisione pseudo-casuale per ciascun caso, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.

Importazione guidata di testo: Passo 3 (File a larghezza fissa)

Figura 3-18
Importazione guidata di testo: Passo 3 (per file a larghezza fissa)



Nel passo 3 vengono fornite informazioni sui casi. I casi sono simili ai record di un database. Ad esempio, ogni persona che risponde a un questionario rappresenta un caso.

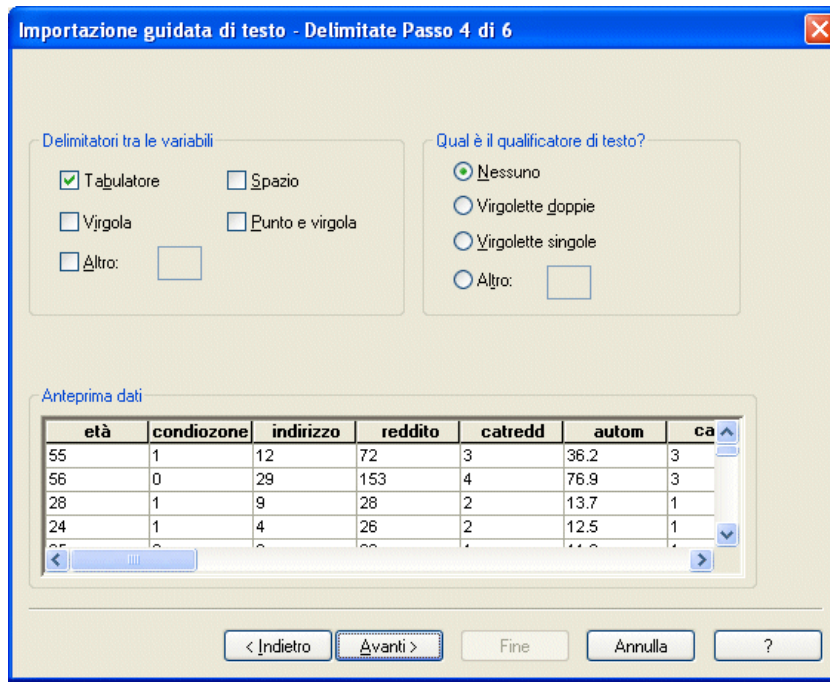
Numero di riga a cui inizia il primo caso di dati. Indica la prima riga del file di dati che contiene valori. Se la prima riga o le prime righe del file di dati contengono etichette descrittive o testo di altro tipo che non rappresenta un valore, questa *non*sarà la riga 1.

N. di righe che rappresentano un caso. Controlla il modo in cui Importazione guidata di testo determina la fine di ogni caso e l'inizio di quello successivo. Ogni variabile è definita dal relativo numero di riga all'interno del caso e dalla posizione di colonna. È necessario specificare il numero di righe per ogni caso per poter leggere correttamente i dati.

Numero di casi da importare. È possibile importare tutti i casi del file di dati, solo i primi *n* casi (*n* (n è un numero specificato dall'utente) o un campione casuale determinato in base a una percentuale specificata. Poiché viene eseguito un processo indipendente di decisione pseudo-casuale per ciascun caso, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.

Importazione guidata di testo: Passo 4 (File delimitati)

Figura 3-19
Importazione guidata di testo: Passo 4 (per file delimitati)



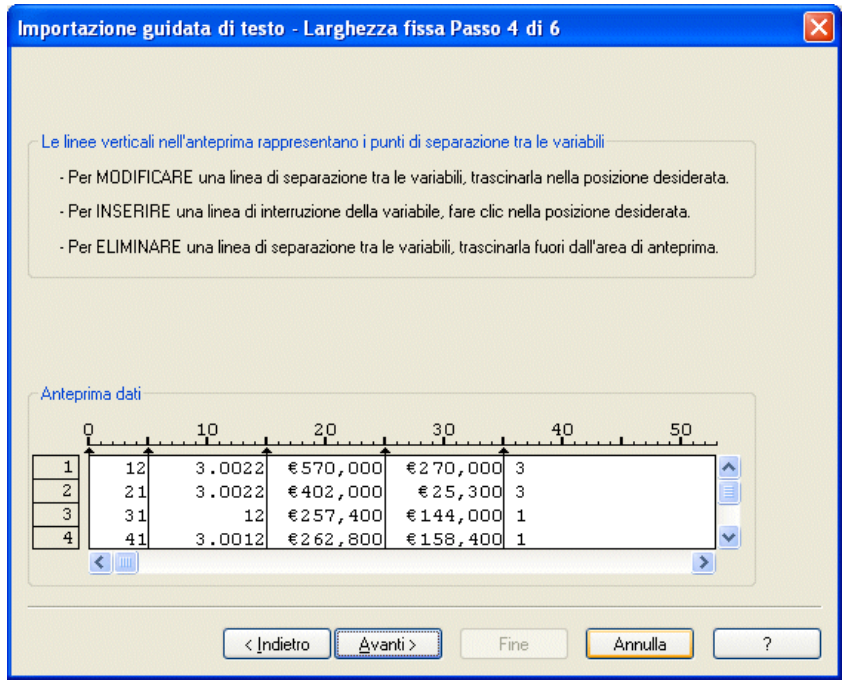
Questo passo illustra la soluzione più opportuna usata da Importazione guidata di testo per leggere un file di dati e consente di modificare il metodo impiegato per leggere le variabili dal file di dati.

Delimitatori tra le variabili. Indica i caratteri o i simboli che separano i valori. È possibile scegliere qualsiasi combinazione di spazi, virgole, punti e virgola, tabulazioni o caratteri di altro tipo. Se sono presenti più delimitatori consecutivi, questi sono considerati valori mancanti.

Qual è il qualificatore di testo? Caratteri usati per includere valori contenenti caratteri delimitatori. Ad esempio, se il delimitatore è costituito da una virgola, i valori che contengono virgole verranno letti in modo inesatto, a meno che non esista un qualificatore di testo contenente il valore e che impedisca alle virgole presenti nel valore di essere interpretate come delimitatori. I file dati in formato CSV esportati da Excel utilizzano le virgolette doppie (“) come qualificatore di testo. Il qualificatore di testo viene visualizzato sia all’inizio sia alla fine del valore, racchiudendo in tal modo l’intero valore.

Importazione guidata di testo: Passo 4 (File a larghezza fissa)

Figura 3-20
Importazione guidata di testo: Passo 4 (per file a larghezza fissa)


Importazione guidata di testo - Larghezza fissa Passo 4 di 6
Le linee verticali nell'anteprima rappresentano i punti di separazione tra le variabili
- Per MODIFICARE una linea di separazione tra le variabili, trascinarla nella posizione desiderata.
- Per INSERIRE una linea di interruzione della variabile, fare clic nella posizione desiderata.
- Per ELIMINARE una linea di separazione tra le variabili, trascinarla fuori dall'area di anteprima.
Anteprima dati
0
10
20
30
40
50
1 12 3.0022 €570,000 €270,000 3
2 21 3.0022 €402,000 €25,300 3
3 31 12 €257,400 €144,000 1
4 41 3.0012 €262,800 €158,400 1
< Indietro
Avanti >
Fine
Annulla
?

Questo passo illustra la soluzione più opportuna usata da Importazione guidata di testo per leggere un file di dati e consente di modificare il metodo impiegato per leggere le variabili dal file di dati. Le linee verticali nella finestra di anteprima indicano le posizioni in cui la procedura Importazione guidata di testo ipotizza possa iniziare una variabile.

Inserire, spostare ed eliminare le linee di interruzione per separare le variabili in base alle necessità. Se per ogni caso vengono usate più righe, selezionarle nell'elenco a discesa e modificare le linee di interruzione delle variabili in base alle necessità.

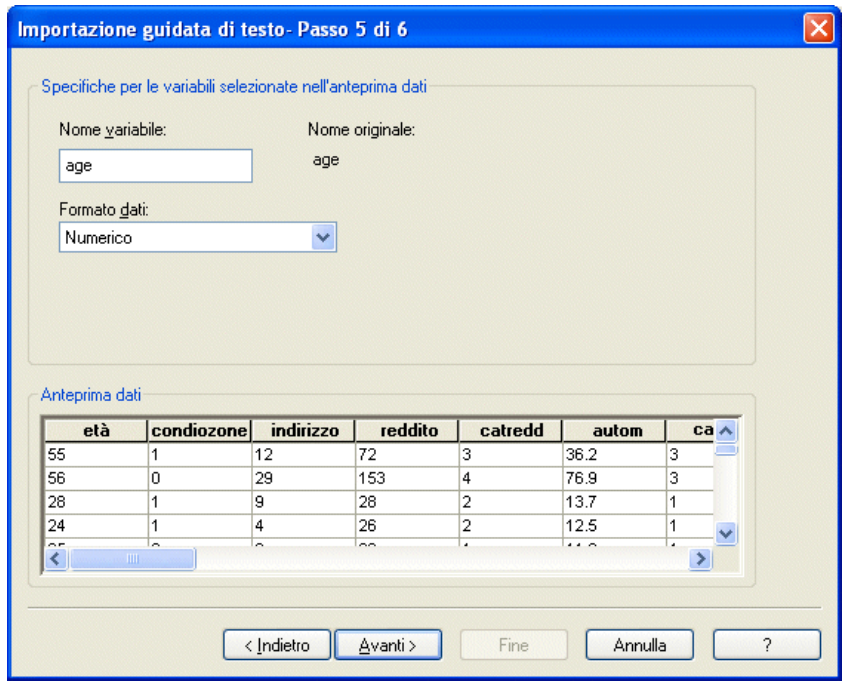
Note:

Per i file di dati generati dal computer che producono un flusso continuo di valori non intercalati da spazi o altri caratteri di separazione, potrebbe essere difficile determinare l'inizio di ogni variabile. Questi file in genere si basano su file di definizione dei dati o su qualche altra descrizione scritta che indica la posizione di riga e colonna per ogni variabile.

L'Importazione guidata di testo viene utilizzata con un carattere proporzionale (a spaziatura fissa) per visualizzare il contenuto dei file. Per i file con carattere non proporzionale, il contenuto del file non risulterà allineato correttamente. È possibile definire questa impostazione del carattere utilizzando l'impostazione Carattere per l'output testuale disponibile nella scheda Viewer della finestra di dialogo Opzioni (menu Modifica).

Importazione guidata di testo: Passo 5

Figura 3-21
Importazione guidata di testo: Passo 5



Questi passi controllano il nome della variabile e il formato dei dati usati da Importazione guidata di testo per leggere le variabili e quali variabili dovranno essere incluse nel file di dati finale.

Nome variabile. È possibile sovrascrivere i nomi delle variabili di default con nomi personalizzati. Se i nomi delle variabili vengono letti dal file di dati, Importazione guidata di testo modificherà automaticamente i nomi delle variabili non conformi alle regole di denominazione delle variabili. Selezionare una variabile nella finestra di anteprima e specificarne il nome.

Formato dati. Selezionare una variabile dalla finestra di anteprima e quindi un formato dall'elenco a discesa. Fare clic tenendo premuto Maiusc per selezionare più variabili contigue oppure fare clic tenendo premuto Ctrl per selezionare più variabili non contigue.

Opzioni di formattazione di Importazione guidata di testo

Le opzioni di formattazione per la lettura delle variabili con Importazione guidata di testo includono:

Non importare. Omette le variabili selezionate dal file di dati importato.

Numerica. I valori validi possono essere numeri, un segno più (+) o meno (-) iniziale e un separatore decimale.

Stringa. I valori validi sono rappresentati da qualsiasi carattere incluso nella tastiera e dagli spazi interposti. Per i file delimitati, è possibile specificare il numero di caratteri per valore, fino a un massimo di 32.767. Per default, Importazione guidata di testo imposta il numero di caratteri in base al valore della stringa più lunga tra le variabili selezionate. Per i file a lunghezza fissa, il numero di caratteri nei valori stringa è determinato dalla posizione delle linee di interruzione delle variabili come definite nel passo 4.

Data/ora. I valori ammessi includono le date nel formato generale *gg-mm-aaaa*, *mm/gg/aaaa*, *gg.mm.aaaa*, *aaaa/mm/gg*, *hh:mm:ss*, nonché in molti altri formati. I mesi possono essere rappresentati in cifre, numeri romani, abbreviazioni di tre lettere o espressi con il nome completo. Selezionare un formato di data dall'elenco.

Dollaro. I valori validi sono numeri con un segno di dollaro iniziale (facoltativo) e la virgola (facoltativa) come separatore delle migliaia.

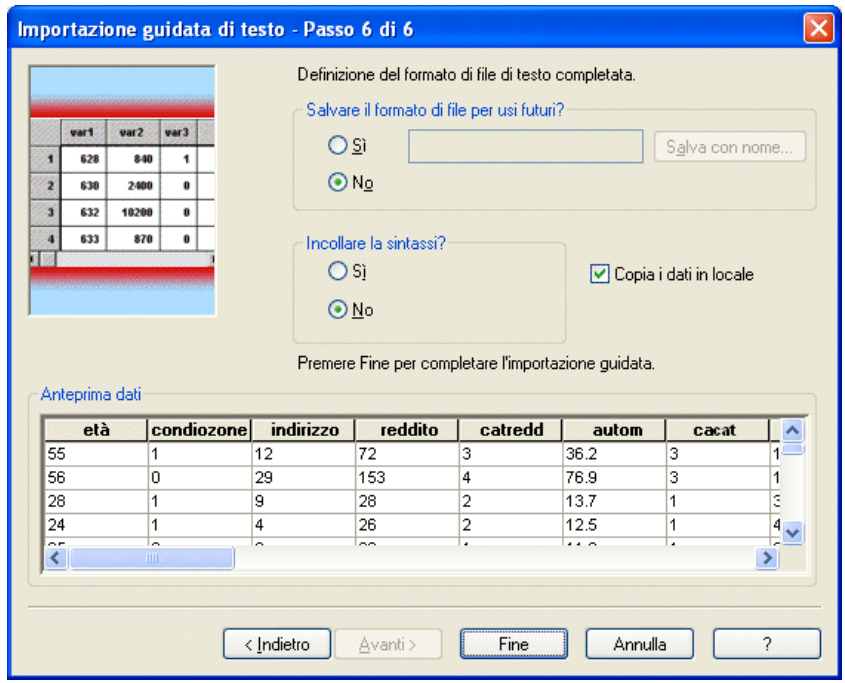
Virgola. Sono ammessi i numeri che usano un punto come separatore dei decimali e la virgola come separatore delle migliaia.

Punto. Sono ammessi i numeri che usano una virgola come separatore dei decimali e il punto come separatore delle migliaia.

Nota: I valori che contengono caratteri incompatibili con il formato selezionato saranno considerati valori mancanti. I valori che contengono uno qualsiasi dei separatori specificati verranno considerati valori multipli.

Importazione guidata di testo: Passo 6

Figura 3-22
Importazione guidata di testo: Passo 6



È il passo conclusivo della procedura. È possibile salvare le impostazioni in un file da utilizzare per l'importazione di file di dati di testo simili. È inoltre possibile incollare nella finestra appropriata la sintassi generata da Importazione guidata di testo. Infine è possibile personalizzare e/o salvare la sintassi in modo da poterla utilizzare in altre sessioni o durante i processi di produzione.

Copia i dati in locale. Crea una copia completa dei dati nella cartella dei file temporanei. La copia dei dati può migliorare le prestazioni.

Lettura di dati Dimensions

È possibile leggere dati da prodotti SPSS Dimensions, compresi Quanvert, Quancept e mrInterview.

Per leggere le sorgenti di dati Dimensions, nel computer in cui viene eseguito SPSS devono essere installati i seguenti elementi:

- .NET framework
- Dimension Data Model e OLE DB Access

Le versioni di tali componenti, compatibili con la presente versione di SPSS, possono essere installate dal CD di installazione SPSS e sono disponibili nel menu di esecuzione automatica. È possibile leggere le sorgenti di dati Dimensions solo in modalità di analisi locale. Questa funzionalità non risulta disponibile in modalità di analisi distribuita quando si utilizza SPSS Server.

Per leggere i dati da una sorgente di dati Dimensions:

▶ Dai menu di qualsiasi finestra SPSS aperta, scegliere:
File
Apri dati Dimensions

▶ Nella scheda Connessione della finestra di dialogo Proprietà di Data Link, specificare i file di metadati, il tipo di dati del caso e il file di dati del caso.

▶ Fare clic su OK.

▶ Nella finestra di dialogo Importazione dati Dimensions selezionare le variabili che si desidera includere e tutti i criteri di selezione del caso.

▶ Fare clic su OK per leggere i dati.

Scheda Connessione della finestra di dialogo Proprietà di Data Link

Per leggere una sorgente di dati Dimensions, è necessario specificare:

Posizione dei metadati. File documento dei metadati (*.mdd*) contenente informazioni sulla definizione del questionario.

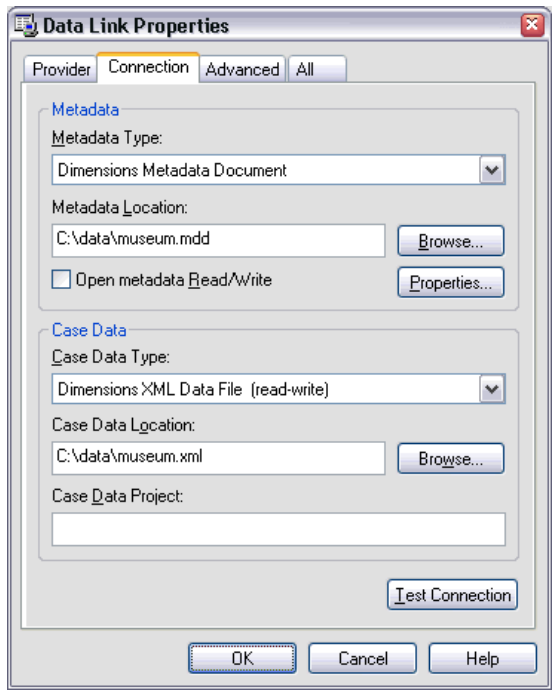
Tipo di dati del caso. Formato del file di dati del caso. I formati disponibili includono:

- File di dati Quancept (DRS). Dati del caso in un file Quancept *.drs*, *.drz* o *.dru*.
- Database Quanvert. Dati del caso in un database Quanvert.

- Database Dimensions (MS SQL Server). Dati del caso in un database relazionale MR SPSS in SQL Server. Questa opzione può essere utilizzata per leggere i dati raccolti con mrInterview.
- File di dati XML Dimensions Dati del caso in un file XML.

Posizione dei dati del caso. File che contiene i dati del caso. Il formato di questo file deve essere coerente con il tipo di dati del caso selezionato.

Figura 3-23
Proprietà di Data Link: scheda Connessione



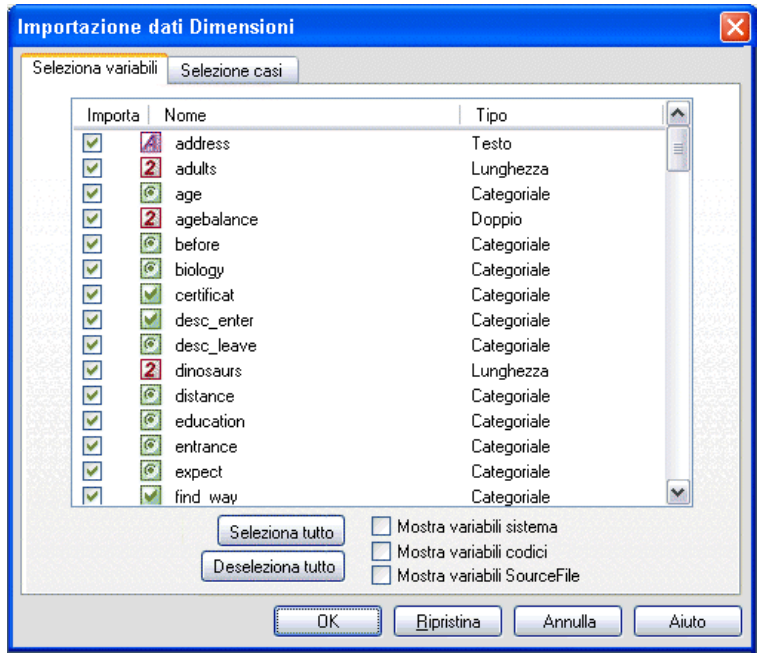
Nota: la misura in cui altre impostazioni presenti nella scheda Connessione, o qualsiasi ulteriore impostazione disponibile in altre schede della finestra di dialogo Proprietà di Data Link, possono, o meno, influire sulla lettura dei dati Dimensions in SPSS non è nota. Di conseguenza è consigliabile evitare di apportare modifiche.

Scheda Seleziona variabili

È possibile selezionare un sottoinsieme di variabili da leggere. Per default, vengono visualizzate e selezionate tutte le variabili standard presenti nella sorgente di dati.

- **Mostra variabili di sistema.** Consente di visualizzare tutte le variabili di "sistema", incluse le variabili che indicano lo stato del colloquio (*in corso*, *completato*, *data di completamento* e così via). A questo punto, è possibile selezionare qualsiasi variabile di sistema che si desidera includere. Per default, vengono escluse tutte le variabili di sistema.
- **Mostra variabili dei codici.** Consente di visualizzare tutte le variabili che rappresentano i codici utilizzati per le risposte aperte "Altro" per le variabili categoriali. A questo punto, è possibile selezionare qualsiasi variabile dei codici che si desidera includere. Per default, vengono escluse tutte le variabili dei codici.
- **Mostra variabili SourceFile.** Consente di visualizzare tutte le variabili che contengono i nomi di file delle immagini delle risposte esaminate. A questo punto, è possibile selezionare qualsiasi variabile SourceFile che si desidera includere. Per default, vengono escluse tutte le variabili SourceFile.

Figura 3-24
Importazione dati Dimensions: scheda Seleziona variabili



Scheda Selezione casi

Per le sorgenti di dati Dimensions che contengono variabili di sistema, è possibile selezionare i casi in base a numerosi criteri di tali variabili. Non è necessario includere le variabili di sistema corrispondenti nell'elenco delle variabili da leggere, ma le variabili di sistema richieste devono risultare presenti nei dati di sorgente per poter applicare i criteri di selezione. In caso contrario, i criteri di selezione corrispondenti verranno ignorati.

Stato della raccolta di dati. È possibile selezionare i dati del rispondente, i dati di test o entrambi. È inoltre possibile selezionare i casi in base a una qualsiasi combinazione dei seguenti parametri relativi allo stato dei colloqui:

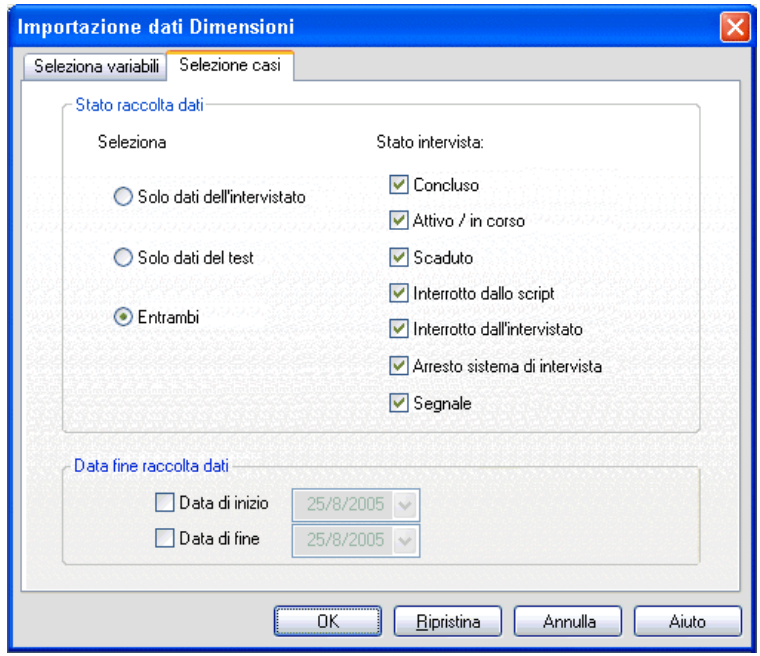
- Correttamente completato
- Attivo/in corso

- Scaduto
- Interrotto dallo script
- Interrotto dal rispondente
- Arresto del sistema di colloquio
- Segnale (terminato da un'istruzione di segnale nello script)

Data di completamento della raccolta dei dati. È possibile selezionare i casi in base alla data di completamento della raccolta dei dati.

- **Data iniziale.** Vengono inclusi i casi in cui la raccolta dei dati risulta terminata alla data specificata o successivamente.
- **Data finale.** Vengono inclusi i casi in cui la raccolta dei dati risulta terminata prima della data specificata. Questo *non* comprende i casi in cui la raccolta dei dati viene completata in corrispondenza della data finale.
- Se si specifica sia una data iniziale che una finale, viene definito un intervallo di date di completamento a partire dalla data iniziale e fino a quella finale (non inclusa).

Figura 3-25
Importazione dati Dimensions: scheda Selezione casi



Informazioni sul file

Un file di dati non contiene semplicemente dati. Contiene anche informazioni sulla definizione delle variabili, ad esempio:

- Nomi delle variabili
- Formati delle variabili
- Etichette descrittive delle variabili e dei valori

Queste informazioni vengono memorizzate nella parte del file di dati riservata al dizionario. Nell'Editor dei dati è possibile visualizzare le informazioni sulla definizione delle variabili. È possibile anche visualizzare le informazioni complete sul dizionario per il file di dati attivo o qualsiasi altro file di dati.

Per ottenere informazioni su un file di dati

- Dai menu nella finestra dell'Editor dei dati, scegliere:
  File
  Per ottenere informazioni su un file di dati
- Per il file di dati attualmente aperto, scegliere File di lavoro.
- Per gli altri file dati, selezionare File esterno, quindi scegliere il file dati.

  Le informazioni sul file di dati verranno visualizzate nel Viewer.

Salvataggio di file di dati

Qualsiasi modifica apportata a un file di dati sarà valida solo fino al termine della sessione corrente di SPSS, a meno che le modifiche non vengano salvate esplicitamente.

Salvataggio di file di dati modificati

- Attivare la finestra dell'Editor dei dati facendo clic in un punto qualsiasi al suo interno.
- Dai menu, scegliere:
  File
  Salva

  Il file di dati modificato verrà salvato e la versione precedente verrà sovrascritta.

Salvataggio di file di dati in formato Excel

È possibile salvare i dati in uno dei tre formati di Microsoft Excel. La scelta del formato dipende dalla versione di Excel che verrà utilizzata per aprire i dati. L'applicazione di Excel non può aprire un file di Excel di una versione più recente. Ad esempio, Excel 5.0 non può aprire un documento di Excel 2000. Al contrario, Excel 2000 può aprire con facilità un documento di Excel 5.0.

A differenza dei file SPSS, nei file in formato Excel esistono alcune limitazioni. Queste limitazioni includono:

- I file di Excel esportati non contengono informazioni sulle variabili, quali valori mancanti e etichette delle variabili.

- Quando si esportano file in Excel 97 e versioni successive, è disponibile un'opzione per aggiungere le etichette dei valori anziché i valori.
- Poiché tutti i file di Excel sono limitati a 256 colonne di dati, nel file esportato vengono incluse solo le prime 256 variabili.
- I file di Excel 4.0 e Excel 5.0/95 sono limitati a 16.384 record o righe di dati. I file di Excel 97–2000 possono contenere fino a 65.536 record. Se i dati superano questi limiti, verrà visualizzato un messaggio di avviso e i dati verranno troncati in base alle dimensioni massime consentite da Excel.

Tipi di variabili

Nella tabella che segue viene indicato il tipo di variabile per cui viene mantenuta la corrispondenza tra i dati originali in file SPSS e i dati esportati in file di Excel.

Tipo di variabile SPSS	**Formato di dati Excel**
Numerico	0.00; #,##0.00; ...
Virgola	0.00; #,##0.00; ...
Dollaro	$#,##0_); ...
Data	gg-mm-aaaa
Tempo (Stima di curve)	hh.mm.ss
Stringa	Generale

Salvataggio di file di dati in formato SAS

Quando i dati vengono salvati come file in formato SAS, viene offerta una speciale gestione per vari aspetti di questi dati. Questi aspetti includono:

- Alcuni caratteri, come @, # e $, consentiti nei nomi di variabili SPSS non sono validi nei file in formato SAS. Questi caratteri vengono sostituiti con un carattere di sottolineatura durante l'esportazione dei dati.
- I nomi delle variabili SPSS che contengono caratteri con più byte (come quelli dell'alfabeto giapponese o cinese) vengono convertiti in nomi di variabili nel formato generico *Vnnn*, dove *nnn* rappresenta un numero intero.
- Le etichette delle variabili SPSS contenenti più di 40 caratteri vengono troncate se esportate in un file SAS versione 6.

- Le eventuali etichette delle variabili SPSS vengono connesse alle etichette delle variabili SAS. Se i dati SPSS non includono alcuna etichetta di variabile, il nome della variabile viene connesso all'etichetta della variabile SAS.
- In un file SAS è possibile includere un solo valore mancante di sistema mentre in un file SPSS possono essere inclusi più valori mancanti di sistema, di conseguenza tutti i valori mancanti di sistema in SPSS verranno connessi a un singolo valore mancante di sistema nel file SAS.

Salvataggio delle etichette dei valori

Le etichette dei valori e i valori e associati al file di dati possono essere salvati in un file di sintassi SAS. Ad esempio, quando vengono esportate le etichette dei valori associate al file di dati *Auto.sav*, il file di sintassi generato contiene:

```
 libname library 'd:\spss\' ;

proc format library = library ;
   value ORIGIN /* Paese di origine */
      1 = 'Italia'
      2 = 'Europa'
      3 = 'Giappone' ;
   value CYLINDER /* Numero di cilindri */
      3 = '3 cilindri'
      4 = '4 cilindri'
      5 = '5 cilindri'
      6 = '6 cilindri'
      8 = '8 cilindri'
   value FILTER__ /* cylrec = 1 | cylrec = 2 (FILTER) */
      0 = 'Not Selected'
      1 = 'Selected' ;

proc datasets library = library ;
modify cars;
   format    ORIGIN ORIGIN.;
```

```
   format  CYLINDER CYLINDER.;
   format  FILTER__ FILTER__.;
quit;
```

Questa caratteristica non è disponibile per il file di trasporto SAS.

Tipi di variabili

Nella tabella che segue viene indicato il tipo di variabile per cui viene mantenuta la corrispondenza tra i dati originali in file SPSS e i dati esportati in file SAS.

Tipo di variabile SPSS	Tipo di variabile SAS	Formato dati SAS.
Numerico	Numerico	12
Virgola	Numerico	12
Punto	Numerico	12
Notazione scientifica	Numerico	12
Data	Numerico	(Data) ad esempio, MMGGAA10,...
Data (ora)	Numerico	Orario 18
Dollaro	Numerico	12
Valuta personalizzata	Numerico	12
Stringa	Carattere	$8

Salvataggio di file di dati in formato Stata

- I dati possono essere scritti in formato Stata 5-8 e nei formati Intercooled e SE format (solo versioni 7-8).
- I file di dati salvati nel formato Stata 5 possono essere letti con Stata 4.
- I primi 80 byte delle etichette delle variabili vengono salvati come etichette di variabili Stata.
- Per le variabili numeriche, i primi 80 byte delle etichette dei valori vengono salvate come etichette dei valori Strata. Per le variabili stringa, le etichette dei valori vengono eliminate.

- Per le versioni 7 e 8, i primi 32 byte dei nomi delle variabili in maiuscolo/minuscolo vengono salvate come nomi di variabili Strata. Per le versioni precedenti, i primi 8 byte dei nomi delle variabili vengono salvati come nomi di variabili Strata. Tutti i caratteri diversi da lettere, numeri e trattini di sottolineatura vengono convertiti in trattini di sottolineatura.
- I nomi delle variabili SPSS che contengono caratteri con più byte (come quelli dell'alfabeto giapponese o cinese) vengono convertiti in nomi di variabili nel formato generico *Vnnn*, dove *nnn* rappresenta un numero intero.
- Per le versioni 5-6 e Intercooled versioni 7-8, vengon salvati i primi 80 byte dei valori stringa. Per Stata SE 7-8, vengono salvati i primi 244 dei valori stringa.
- Per le versioni 5-6 e Intercooled versioni 7-8, vengono salvate solo le prime 2.047 variabili. Per Stata SE 7-8, vengono salvate solo le prime 32.767 variabili.

Tipo variabile SPSS	Tipo variabile Stata	Formato di dati Stata
Numerico	Numerico	g
Virgola	Numerico	g
Punto	Numerico	g
Notazione scientifica	Numerico	g
Data, Data/ora	Numerico	D_m_Y
Ora, DOra	Numerico	g (numero di secondi)
Giorno della settimana	Numerico	g (1-7)
Moyr	Numerico	g (1-12)
Dollaro	Numerico	g
Valuta personalizzata	Numerico	g
Stringa	Stringa	s

Salvataggio di un nuovo file di dati SPSS o in formato diverso

▶ Attivare la finestra dell'Editor dei dati facendo clic in un punto qualsiasi al suo interno.

▶ Dai menu, scegliere:

File
  Salva con nome...

▶ Selezionare un tipo di file dall'elenco a discesa.

▶ Specificare un nome per il nuovo file di dati.

Per scrivere i nomi delle variabili nella prima riga di file di dati in formato foglio elettronico o tab delimitato:

▶ Fare clic su Scrivi i nomi di variabile su foglio elettronico nella finestra di dialogo Salva con nome.

Per salvare le etichette dei valori anziché i valori dei dati in formato Excel 97:

▶ Fare clic su Salva etichette di valore se definite anziché valori di dati nella finestra di dialogo Salva con nome.

Per salvare le etichette dei valori in un file di sintassi SAS (disponibile solo se viene selezionato il tipo di file SAS):

▶ Fare clic su Salva etichette di valore in un file .sas nella finestra di dialogo Salva con nome.

Salvataggio di dati: Tipi di file di dati

È possibile salvare i dati nei seguenti formati:

SPSS (*.sav). Formato SPSS.

- I file di dati salvati in formato SPSS non possono essere letti con versioni del prodotto precedenti alla 7.5.
- Quando in SPSS 10.x o 11.x si utilizzano file di dati con nomi di variabili che contengono più di otto byte, vengono utilizzate le versioni a otto byte dei nomi di variabili univoci,— mantenendo tuttavia i nomi originali delle variabili per l'utilizzo nella versione 12.0 o successiva. Nelle versioni di SPSS precedenti alla 10,0, i nomi originali delle variabili più lunghi di quanto consentito andavano perduti se si salvava il file di dati.
- Se si utilizzano file di dati con variabili stringa di lunghezza superiore a 255 byte nelle versioni SPSS precedenti alla 13.0, le variabili stringa vengono divise in più variabili stringhe a 255 byte.

SPSS 7.0 (*.sav). Formato SPSS 7.0 per Windows. I file di dati salvati in formato SPSS 7.0 possono essere letti con SPSS per Windows versione 7.0 e precedenti, ma non includono insiemi a risposta multipla definiti né informazioni sull'inserimento di dati in Windows.

SPSS/PC+ (*.sys). Formato SPSS/PC+. Se il file di dati contiene più di 500 variabili, verranno salvate solo le prime 500. Per le variabili con più valori mancanti definiti dall'utente, i valori mancanti aggiuntivi verranno ricodificati nel primo valore mancante definito dall'utente.

Portabile SPSS (*.por). Formato portatile SPSS che può essere letto da altre versioni di SPSS in altri sistemi operativi (ad esempio, Macintosh o UNIX). I nomi di variabili possono includere fino a otto byte e, se necessario, vengono convertiti automaticamente in nomi univoci a otto byte.

Tab delimitato (*.dat). File di testo ASCII con i valori separati da tabulazioni.

ASCII fisso (*.dat). File di testo ASCII in formato fisso, in cui vengono utilizzati i formati di scrittura predefiniti per tutte le variabili. Tra i campi della variabile non sono presenti tabulazioni né spazi.

Excel 2.1 (*.xls). Foglio elettronico di Microsoft Excel 2.1. Il numero massimo di variabili è 256 e il numero massimo di righe è 16.384.

Excel 97 e versioni successive (*.xls). Foglio elettronico di Microsoft Excel 97/2000/XP. Il numero massimo di variabili è 256 e il numero massimo di righe è 65.536.

1-2-3 versione 3.0 (*.wk3). Foglio elettronico Lotus 1-2-3, versione 3,0. È possibile salvare un numero massimo di 256 variabili.

1-2-3 versione 2.0 (*.wk1). Foglio elettronico Lotus 1-2-3, versione 2.0. È possibile salvare un numero massimo di 256 variabili.

1-2-3 versione 1.0 (*.wks). Foglio elettronico Lotus 1-2-3, versione 1A. È possibile salvare un numero massimo di 256 variabili.

SYLK (*.slk). Formato Symbolic Link per fogli elettronici di Microsoft Excel e Multiplan. È possibile salvare un numero massimo di 256 variabili.

dBASE IV (*.dbf). Formato dBASE IV.

dBASE IV (*.dbf). Formato dBASE IV.

dBASE IV (*.dbf). Formato dBASE IV.

SAS v6 per Windows (*.sd2). Formato di file SAS v6 per Windows/OS2.

SAS v6 per UNIX (*.ssd01). Formato di file SAS v6 per UNIX (Sun, HP, IBM).

SAS v6 per Alpha/OSF (*.ssd04). Formato di file SAS v6 per Alpha/OSF (DEC UNIX).

SAS v7+ Windows estensione corta (*.sd7). Versione SAS 7–8 per Windows, formato con nome di file corto.

SAS v7+ Windows estensione lunga (*.sas7bdat). Versioni SAS 7–8 per Windows, formato con nome di file lungo.

SAS v7 per UNIX (*.ssd01). SAS v8 per UNIX.

Trasporto SAS (*.xpt). File di trasporto SAS.

Stata Versioni 4–5 (*.dta).

Stata Versione 6 (*.dta).

Stata Versione 7 Intercooled (*.dta).

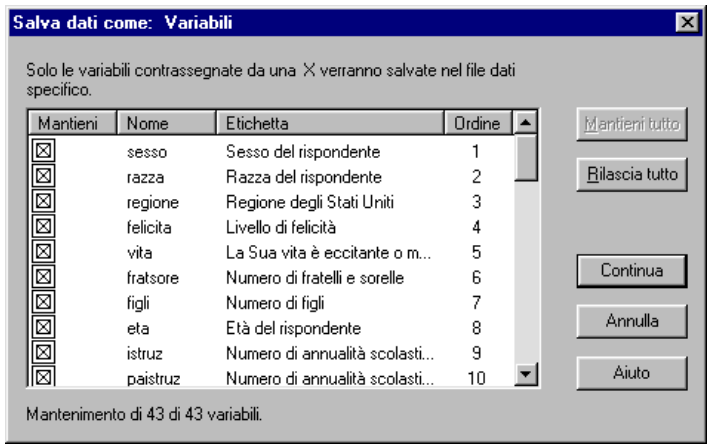
Stata Versione 7 SE (*.dta).

Stata Versione 8 Intercooled (*.dta).

Stata Versione 8 SE (*.dta).

Salvataggio di sottoinsiemi di variabili

Figura 3-26
Finestra di dialogo Salva dati come variabili


Salva dati come: Variabili
Solo le variabili contrassegnate da una X verranno salvate nel file dati specifico.
Mantieni
Nome
Etichetta
Ordine
sesso Sesso del rispondente 1
razza Razza del rispondente 2
regione Regione degli Stati Uniti 3
felicita Livello di felicità 4
vita La Sua vita è eccitante o m... 5
fratsore Numero di fratelli e sorelle 6
figli Numero di figli 7
eta Età del rispondente 8
istruz Numero di annualità scolasti... 9
paistruz Numero di annualità scolasti... 10
Mantieni tutto
Rilascia tutto
Continua
Annulla
Aiuto
Mantenimento di 43 di 43 variabili.

La finestra di dialogo Salva dati come permette di selezionare le variabili da salvare nel nuovo file di dati. Per impostazione predefinita, tutte le variabili verranno salvate. Deselezionare le variabili che non si desidera salvare oppure fare clic su Rilascia tutto e selezionare le variabili da salvare.

Per salvare un sottoinsieme di variabili

- Attivare la finestra dell'Editor dei dati facendo clic in un punto qualsiasi al suo interno.
- Dai menu, scegliere:
  File
  Salva con nome...
- Seleziona i tipi di file.
- Fare clic su Variabili.
- Selezionare le variabili che si desidera salvare.

Opzioni per il salvataggio di file

Nei fogli elettronici e nei file tab delimitati è possibile scrivere il nome delle variabili nella prima riga.

Protezione dei dati originali

Per evitare che i dati originali vengano modificati o eliminati accidentalmente, è possibile contrassegnare il file come di sola lettura.

- Dai menu dell'Editor dei dati, scegliere:
  File
  Contrassegna file come di sola lettura

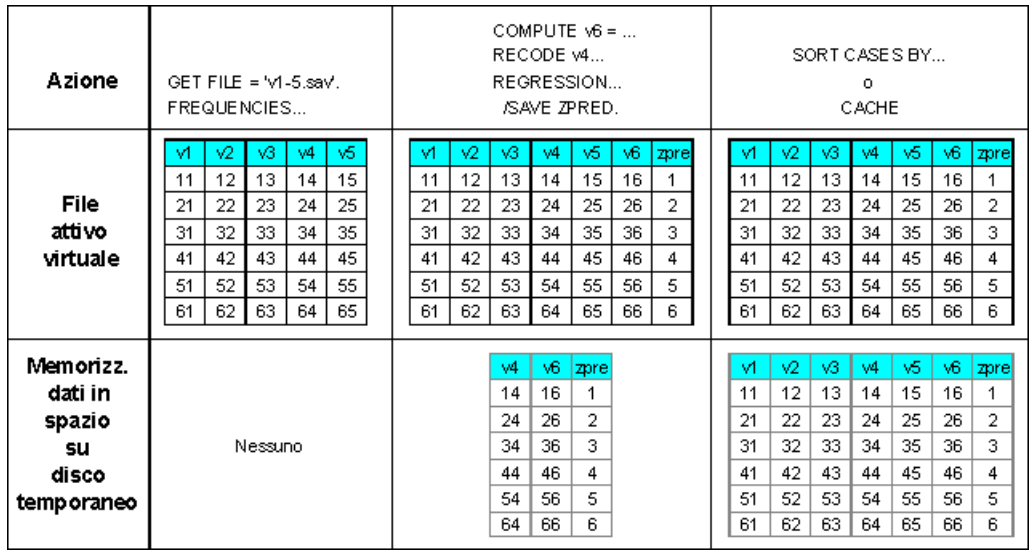
Se in seguito si apportano modifiche ai dati e quindi si prova a salvare il file, sarà possibile salvarlo solo con un nome file diverso. In questo modo, i dati originali non verranno alterati.

È possibile reimpostare i permessi di lettura/scrittura per il file scegliendo Contrassegna file come di scrittura dal menu File.

File attivo virtuale

Il file attivo virtuale consente di lavorare con file di dati di grandi dimensioni senza che sia necessario liberare una quantità uguale o maggiore di spazio su disco temporaneo. Nella maggior parte delle procedure di analisi e rappresentazione grafica, i dati di origine vengono riletti ogni volta che si esegue una procedura diversa. Le procedure che modificano i dati necessitano di una certa quantità di spazio su disco temporaneo per tenere traccia delle modifiche. Inoltre, alcune azioni richiedono sempre una quantità di spazio su disco sufficiente a contenere almeno una copia dell'intero file di dati.

Figura 3-27
Requisiti di spazio su disco temporaneo



Le azioni che non richiedono spazio su disco temporaneo sono:

- Lettura dei file di dati di SPSS
- Unione di due o più file di dati SPSS
- Lettura di tabelle di database con la procedura Creazione guidata di query
- Unione di un file di dati SPSS con una tabella di database
- Esecuzione delle procedure di lettura dei dati (ad esempio, Frequenze, Tavole di contingenza, Esplora)

Le azioni che creano una o più colonne di dati nello spazio su disco temporaneo sono:

- Calcolo di nuove variabili
- Registrazione delle variabili esistenti
- Esecuzione di procedure che creano o modificano variabili (ad esempio, il salvataggio dei valori attesi nella regressione lineare)

Le azioni che creano una copia dell'intero file di dati nello spazio su disco temporaneo sono:

- Lettura di file di Excel
- Esecuzione di procedure che ordinano i dati (ad esempio, Ordina casi, Distingui)
- Lettura dei dati con i comandi `GET TRANSLATE` o `DATA LIST`
- Uso dell'utilità di inserimento dati nella cache o del comando `CACHE`
- Avvio da SPSS di altre applicazioni che leggono il file di dati (ad esempio, AnswerTree, DecisionTime)

Nota: il comando `GET DATA` offre funzionalità simili a quelle offerte dal comando `DATA LIST`, senza però creare una copia dell'intero file di dati nello spazio su disco temporaneo. Il comando `SPLIT FILE` nella sintassi dei comandi non esegue l'ordinamento del file di dati e quindi non crea una copia del file di dati. Questo comando richiede tuttavia i dati ordinati per un funzionamento appropriato e l'interfaccia della finestra di dialogo relativa a questa procedura ordina automaticamente il file di dati, creando una copia dell'intero file di dati. La sintassi dei comandi non è disponibile nella versione per studenti.

Azioni che creano una copia intera del file di dati per default:

- Lettura di database con Creazione guidata di query
- Lettura di file di testo con Importazione guidata di testo

L'Importazione guidata testo fornisce un'impostazione opzionale che consente di memorizzare automaticamente i dati nella cache. Questa opzione è selezionata per default. È possibile disattivarla deselezionando la casella di controllo Copia i dati in locale. Per la Creazione guidata di query, è possibile incollare la sintassi dei comandi generata ed eliminare il comando `CACHE`.

Creazione di una copia dei dati

Benché il file attivo virtuale consenta di ridurre notevolmente la quantità di spazio su disco temporaneo richiesta, l'assenza di una copia temporanea del file "attivo" comporta la rilettura dei dati di origine ogni volta che si esegue una procedura. Per i file di dati di grandi dimensioni letti da una sorgente esterna, la creazione di una copia temporanea dei dati può migliorare le prestazioni. Nel caso di tabelle lette da un database, ciò significa che la query SQL che legge i dati dal database deve essere rieseguita per ogni comando o procedura che richiede la lettura dei dati. Dato che praticamente tutte le procedure di analisi statistica e di rappresentazione grafica richiedono una lettura dei dati, la query SQL viene rieseguita a ogni procedura, provocando un notevole aumento dei tempi di elaborazione quando si esegue un numero significativo di procedure.

Se il computer su cui viene eseguita l'analisi (sia esso un computer locale o un server remoto) dispone di spazio su disco sufficiente, è possibile evitare l'esecuzione continua delle query SQL e ridurre i tempi di elaborazione creando una copia cache dei dati del file attivo. Tale copia è una copia temporanea completa dei dati.

Nota: per default, Creazione guidata di query crea automaticamente una copia di dati ma se si utilizza il comando `GET DATA` nella sintassi dei comandi per leggere un database, tale copia non viene creata automaticamente. La sintassi dei comandi non è disponibile nella versione per studenti.

Per creare una copia dei dati

▶ Dai menu, scegliere:

File
Inserisci dati nella cache...

▶ Fare clic su OK o su In cache ora.

Se si fa clic su OK, la volta successiva che il programma legge i dati (ad esempio, alla successiva esecuzione di una procedura statistica) viene creata una copia dei dati che consente di evitare un ulteriore lettura degli stessi. Se si fa clic su In cache ora, si

ottiene la creazione immediata di una copia dei dati, non necessaria nella maggior parte dei casi. In cache ora è utile soprattutto per due ragioni:

- Una sorgente dati è "bloccata" e non può essere aggiornata da nessuno fino al termine della sessione, all'apertura di un'altra sorgente dati o all'inserimento dei dati nella cache.
- Nel caso di sorgenti dati di grandi dimensioni, lo scorrimento del contenuto della Visualizzazione dati nell'Editor dei dati è molto più veloce se questi vengono inseriti nella cache.

Per eseguire una copia automatica dei dati

È possibile utilizzare il comando `SET` per creare automaticamente una copia dei dati dopo un numero specifico di modifiche nel file di dati attivo. Per default, il file di dati attivo viene copiato automaticamente dopo 20 modifiche in tale file.

▶ Dai menu, scegliere:

File
Nuovo
Sintassi

▶ Nella finestra di sintassi, digitare SET CACHE n. (dove *n* indica il numero di modifiche apportate al file di dati attivo prima di copiarlo nella cache).

▶ Dai menu della finestra di sintassi, scegliere:

Esegui
Tutto

Nota: l'impostazione della cache non è valida in tutte le sessioni. A ogni avvio di una nuova sessione, il valore viene reimpostato al valore predefinito, ovvero 20.

Capitolo
4

Analisi distribuita

La modalità di analisi distribuita consente di utilizzare un computer che non sia quello locale (o desktop) per l'esecuzione di attività che prevedono un uso intensivo della memoria. Poiché i server remoti utilizzati per l'analisi distribuita sono in genere più potenti e veloci del computer locale, l'analisi distribuita può ridurre significativamente i tempi di elaborazione. L'analisi distribuita con un server remoto può rivelarsi vantaggiosa quando l'elaborazione prevede:

- L'utilizzo di file di dati di grandi dimensioni, in particolare dati letti da database.
- Attività che richiedono un uso intensivo della memoria. Tutte le attività che presuppongono tempi di elaborazione lunghi in modalità di analisi locale possono essere eseguite più velocemente in modalità di analisi distribuita.

L'analisi distribuita è efficace solo nel caso di attività connesse ai dati, come lettura e trasformazione di dati, elaborazione di nuove variabili e calcolo di statistiche, mentre non lo è nel caso di attività connesse alla modifica dell'output, quali manipolazione di tabelle pivot o modifica di grafici.

Nota: l'analisi distribuita può essere utilizzata solo se si dispone di una versione locale del software e si ha accesso a una versione server con licenza dello stesso software installato su un server remoto.

Confronto tra analisi distributa e locale

Di seguito vengono riportate alcune indicazioni di carattere generale utili per scegliere la modalità di analisi locale o distribuita:

Accesso ai database. Le sessioni che prevedono l'esecuzione di query a database vengono eseguite in modo più rapido in modalità distribuita se il server dispone di un migliore accesso al database oppure se è in esecuzione sulla stessa macchina del motore del database. Inoltre, se il programma di accesso al database necessario è

disponibile solo sul server, oppure se l'amministratore della rete vieta di scaricare tabelle di dati di grandi dimensioni, sarà possibile accedere al database solo in modalità distribuita.

Rapporto tra calcoli e output. I comandi che eseguono una grande quantità di calcoli e producono output di piccole dimensioni (ad esempio, poche tabelle pivot e di piccole dimensioni, testi brevi, pochi o semplici grafici), sono i più appropriati per l'esecuzione in modalità distribuita. Il livello di miglioramento dipende in larga parte dalla potenza di calcolo del server remoto.

Sessioni poco impegnative. Le sessioni eseguibili rapidamente in modalità locale vengono sempre eseguite più lentamente in modalità distribuita, a causa dell'overhead client/server inerente.

Grafici. I grafici orientati ai casi, come grafici a dispersione, grafici dei residui delle regressioni e grafici sequenza, richiedono la presenza di dati non elaborati sul computer locale. Nel caso di file di dati o di tabelle di database di grandi dimensioni, è possibile che si abbia un peggioramento delle prestazioni in modalità distribuita in quanto i dati devono essere inviati dal server remoto al computer locale. Gli altri grafici sono basati su dati riepilogati o aggregati e vengono elaborati a velocità adeguate, poiché l'aggregazione dei dati viene eseguita dal server.

Grafici interattivi. Se i dati grezzi vengono salvati con grafica interattiva (impostazione opzionale), è possibile che vengano trasferite grandi quantità di dati dal server remoto al computer con un conseguente incremento del tempo necessario per salvare i risultati.

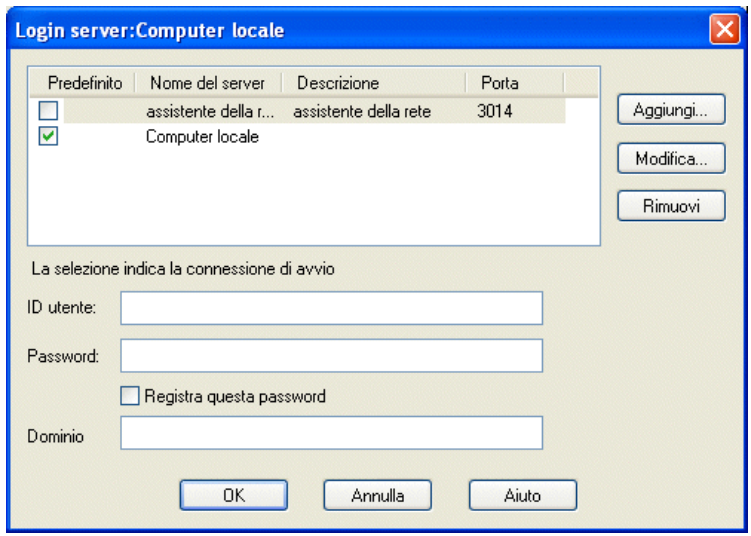
Tabelle pivot. La creazione di tabelle pivot di notevoli dimensioni può risultare più lenta in modalità distribuita. Ciò è particolarmente vero per la procedura Cubi OLAP e per le tabelle che contengono dati di singoli casi, come quelli disponibili nella procedura Riassumi.

L'output testuale prodotto da alcune procedure. Maggiore è la quantità di testo prodotta, più lenta risulta l'elaborazione in modalità distribuita, poiché il testo viene prodotto sul server remoto e quindi copiato sul computer locale per essere visualizzato. Il testo ha tuttavia un overhead basso e la trasmissione risulta spesso rapida.

Login al server

La finestra di dialogo Login server consente di selezionare il computer che elabora i comandi ed esegue le procedure, È possibile selezionare il computer locale o un server remoto.

Figura 4-1
Finestra di dialogo Login server



È possibile aggiungere, modificare o eliminare server remoti dall'elenco. I server remoti richiedono in genere un ID utente e una password, ma in certi casi anche il nome di dominio. Per informazioni relative a server disponibili, ID utente e password, nomi di dominio e altre informazioni per la connessione, rivolgersi all'amministratore di sistema.

È possibile selezionare un server predefinito e memorizzare ID utente, nome di dominio e password associati a qualsiasi server. All'avvio di una nuova sessione, viene eseguita automaticamente la connessione al server predefinito.

Aggiunta e modifica delle impostazioni di login al server

Utilizzare la finestra di dialogo Impostazioni login al server per aggiungere o modificare le informazioni per la connessione ai server remoti da utilizzare in modalità di analisi distribuita.

Figura 4-2
Finestra di dialogo Impostazioni login al server

Per un elenco dei server disponibili, dei numeri di porta per i server e altre informazioni per la connessione, rivolgersi all'amministratore del sistema. Non utilizzare Secure Sockets Layer a meno che non venga richiesto specificatamente dall'amministratore.

Nome server. Il "nome" di un server può essere un nome alfanumerico assegnato a un computer (ad esempio, NetworkServer) o un indirizzo IP univoco assegnato a un computer (ad esempio, 202.123.456.78).

Numero porta. Il numero di porta corrisponde alla porta che il software del server utilizza per le comunicazioni.

Descrizione. È possibile immettere una descrizione facoltativa da visualizzare nell'elenco dei server.

Connetti con Secure Socket Layer. Secure Socket Layer (SSL) consente di crittografare le richieste di analisi distribuita durante l'invio al server SPSS remoto. Prima di utilizzare SSL, contattare l'amministratore di sistema. Per attivare questa opzione, è necessario configurare SSL sul computer desktop e sul server.

Per selezionare, cambiare o aggiungere server

▶ Dai menu, scegliere:

File
Cambia server...

Per selezionare un server predefinito:

▶ Nell'elenco dei server selezionare la casella accanto al server che si desidera utilizzare.

▶ Immettere l'ID utente, il nome di dominio e la password forniti dall'amministratore.

Nota: All'avvio di una nuova sessione, viene eseguita automaticamente la connessione al server predefinito.

Per cambiare server:

▶ Selezionare il server dall'elenco.

▶ Se necessario, immettere l'ID utente, il nome di dominio e la password.

Nota: Se durante una sessione si cambia server, tutte le finestre aperte vengono chiuse. Prima dell'effettiva chiusura delle finestre, viene chiesto se si desidera salvare le modifiche.

Per aggiungere un server:

▶ Rivolgersi all'amministratore per le informazioni di connessione al server.

▶ Fare clic su Aggiungi per aprire la finestra di dialogo Impostazioni login al server.

▶ Immettere le informazioni di connessione e le impostazioni opzionali e fare clic su OK.

Per modificare un server:

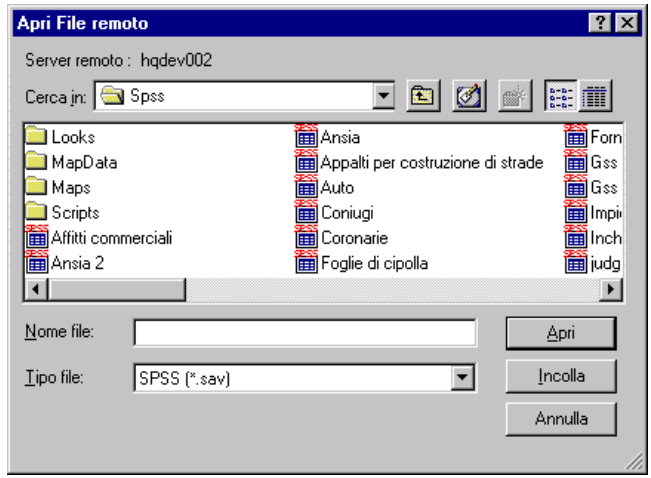
▶ Rivolgersi all'amministratore per le informazioni di connessione al server aggiornate.

▶ Fare clic su Modifica per aprire la finestra di dialogo Impostazioni login al server.

▶ Apportare le modifiche e fare clic su OK.

Apertura di file di dati da un server remoto

Figura 4-3
Finestra di dialogo Apri file remoto



Nell'analisi distribuita la finestra di dialogo Apri file remoto sostituisce la finestra di dialogo Apri file standard.

- Il contenuto dell'elenco di file, cartelle e unità disponibili varia a seconda di quanto risulta disponibile sul server remoto. Il nome del server corrente è indicato nella parte superiore della finestra di dialogo.
- In modalità di analisi distribuita, non sarà possibile accedere ai file di dati del computer locale se non si imposta la condivisione dell'unità e delle cartelle contenenti quei file.
- Se il server esegue un sistema operativo diverso (ad esempio, sul proprio computer è installato Windows e il server esegue UNIX), è probabile che non si possa accedere ai file di dati locali in modalità di analisi distribuita, anche se si trovano in cartelle condivise.

Per aprire i file di dati da un server remoto

▶ Eseguire il login al server remoto se non lo si è fatto in precedenza.

- In base al tipo di file di dati da aprire, dai menu, scegliere:

File
  Apri
    Dati...

oppure

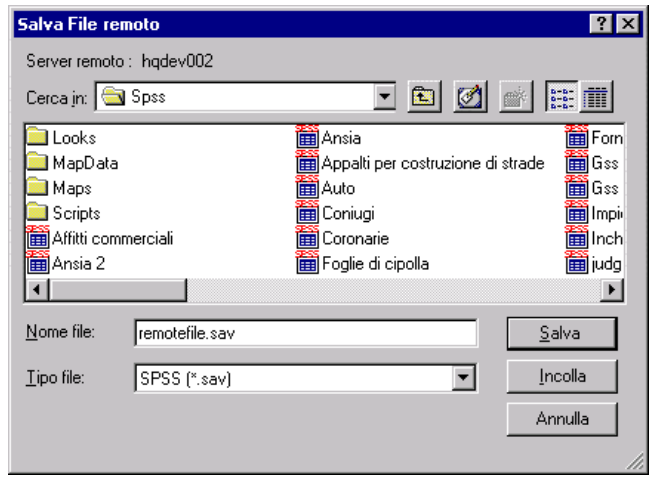
File
  Apri database

oppure

File
  Leggi dati testuali...

Salvataggio di file di dati da un server remoto

Figura 4-4
Finestra di dialogo Salva file remoto



In modalità di analisi distribuita, la finestra di dialogo Salva file remoto sostituisce la finestra di dialogo Salva file standard.

Il contenuto dell'elenco delle cartelle e unità disponibili varia a seconda di quanto risulta disponibile sul server remoto. Il nome del server corrente è indicato nella parte superiore della finestra di dialogo. Non sarà possibile accedere alle cartelle del computer locale se non si imposta la condivisione dell'unità e delle cartelle. Se il server esegue un sistema operativo diverso (ad esempio, sul proprio computer è installato Windows e il server esegue UNIX), è probabile che non si possa accedere

ai file di dati locali in modalità di analisi distribuita, anche se si trovano in cartelle condivise. Per salvare i file di dati in una cartella locale è necessario che i permessi per le cartelle condivise includano il permesso di scrittura nella cartella.

Per salvare i file di dati da un server remoto

- Attivare la finestra dell'Editor dei dati.
- Dai menu, scegliere:
  File
  Salva (o Salva con nome...)

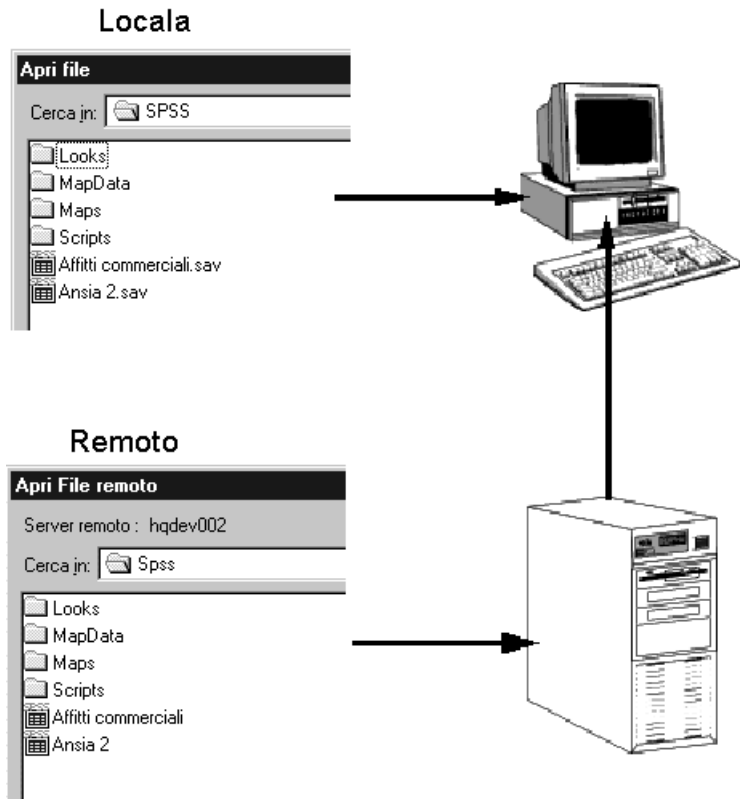
Accesso ai file di dati in modalità di analisi locale e distribuita

La modalità di visualizzazione dei file di dati, delle cartelle (directory) e delle unità nel computer locale e nella rete dipende dal computer in cui vengono eseguite l'elaborazione dei comandi e le procedure, —che non è necessariamente il computer locale.

Modalità di analisi locale. Quando si utilizza il computer locale come "server", la modalità di visualizzazione di file di dati, cartelle e unità nella finestra di dialogo di accesso ai file (per l'apertura dei file di dati) è simile a quella di altre applicazioni o di Esplora risorse di Windows. È possibile visualizzare tutti i file di dati e le cartelle del computer, nonché i file e le cartelle delle unità di rete installate e in genere visualizzate.

Modalità di analisi distribuita. Quando si utilizza un altro computer come "server remoto" per l'esecuzione di comandi e procedure, i file di dati, le cartelle e le unità vengono visualizzati secondo la prospettiva del server remoto. Sebbene i nomi delle cartelle e delle unità come *Programmi* e *C* possano sembrare familiari, questi *non* si riferiscono alle cartelle e alle unità presenti sul computer locale; ma alle cartelle e alle unità del server remoto.

Figura 4-5
Visualizzazioni locali e remote



In modalità di analisi distribuita, non sarà possibile accedere ai file di dati del computer locale se non si imposta la condivisione dell'unità e delle cartelle contenenti quei file. Se il server esegue un sistema operativo diverso (ad esempio, sul proprio computer è installato Windows e il server esegue UNIX), è probabile che non si possa accedere ai file di dati locali in modalità di analisi distribuita, anche se si trovano in cartelle condivise.

La modalità di analisi distribuita non equivale all'accesso ai file di dati che risiedono su un altro computer della rete. È possibile accedere ai file di dati presenti su altre periferiche di rete sia in modalità di analisi locale sia in modalità di analisi distribuita. In modalità locale è possibile accedere ad altre periferiche di rete dal computer locale. In modalità distribuita vi si accede dal server remoto.

Se non si è certi della modalità di analisi in uso, leggere la barra del titolo della finestra di dialogo per l'accesso ai file di dati. Se la barra del titolo contiene la parola *Remoto* (come in Apri file remoto) oppure se nella parte superiore della finestra di dialogo compare la dicitura Server remoto: [nome server], si sta utilizzando la modalità di analisi distribuita.

Nota: Ciò vale solo per le finestre di dialogo utilizzate per l'accesso ai file di dati (ad esempio, Apri dati, Salva dati, Apri database e Applica dizionario dati). Per tutti gli altri tipi (ad esempio, file Viewer, file di sintassi e file script), si utilizza sempre la visualizzazione locale.

Per impostare i permessi di condivisione per un'unità o una cartella

- In Risorse del computer fare clic sulla directory o sull'unità da condividere.
- Nel menu File fare clic su Proprietà.
- Fare clic sulla scheda Condivisione e poi su Condividi con nome.

Per ulteriori informazioni sulla condivisione di unità e cartelle, consultare la Guida in linea del sistema operativo.

Procedure disponibili in modalità di analisi distribuita

In modalità di analisi distribuita, sono disponibili solo le procedure installate sia nella versione locale che nella versione del server remoto.

Se sul computer locale sono stati installati componenti non disponibili sul server remoto e si passa dal computer locale al server remoto, le procedure in questione verranno rimosse dai menu. Quindi, le sintassi di comando corrispondenti genereranno errori. Se si torna alla modalità locale, tutte le procedure interessate vengono ripristinate.

Utilizzo di specifiche di percorso UNC

In modalità di analisi distribuita, con la versione del server di Windows NT per SPSS, le specifiche di percorso relativo per i file di dati si riferiscono al server corrente e non al computer locale. Ciò significa che un percorso quale *c:\mydocs\mydata.sav*) non punta verso una directory e un file presenti sull'unità c locale, ma verso una directory e un file del disco rigido del server remoto. Se la directory e/o il file non

esistono sul server remoto, viene generato un errore nella sintassi del comando come mostra il seguente esempio:

```
GET FILE='c:\documenti\dati.sav'.
```

Se si sta eseguendo la versione del server Windows NT per SPSS, è possibile utilizzare le specifiche UNC (Universal Naming Convention, convenzione universale di denominazione) quando si accede ai file di dati con la sintassi dei comandi. La forma generica di una specifica UNC è la seguente:

```
\\nomeserver\condivisione\percorso\nomefile
```

- *Nomeserver* è il nome del computer in cui si trova il file di dati.
- *Condivisione* è la cartella (directory) su quel computer designata come cartella condivisa.
- *Percorso* è il percorso delle eventuali sottocartelle (sottodirectory) della cartella condivisa.
- *Nomefile* è il nome del file di dati.

Esempio:

```
GET FILE = '\\hqdev01\public\luglio\vendite.sav'.
```

Se al computer non è stato assegnato un nome, è possibile utilizzare un indirizzo IP, ad esempio:

```
GET FILE='\\204.125.125.53\public\luglio\vendite.sav'.
```

Anche con le specifiche di percorso UNC, è possibile accedere ai file di dati presenti unicamente nelle periferiche e nelle cartelle condivise. Quando si utilizza la modalità di analisi distribuita, ciò vale anche per i file di dati presenti sul computer locale.

Server UNIX. Nelle piattaforme UNIX, non esiste un equivalente del percorso UNC e tutti i percorsi alle directory devono essere percorsi assoluti che partono dalla directory principale del server. I percorsi relativi non sono consentiti. Ad esempio, se il file di dati si trova in */bin/spss/data* e anche la directory corrente è */bin/spss/data*, `GET FILE='vendite.sav'` non è valido in quanto è necessario specificare tutto il percorso come segue:

```
GET FILE='/bin/spss/data/sales.sav'.
```

Capitolo
5

Editor dei dati

L'Editor dei dati è uno strumento utile per la creazione e la modifica di file di dati, analogo a quello dei fogli di calcolo. La finestra dell'Editor dei dati viene aperta automaticamente all'apertura di ogni sessione di SPSS.

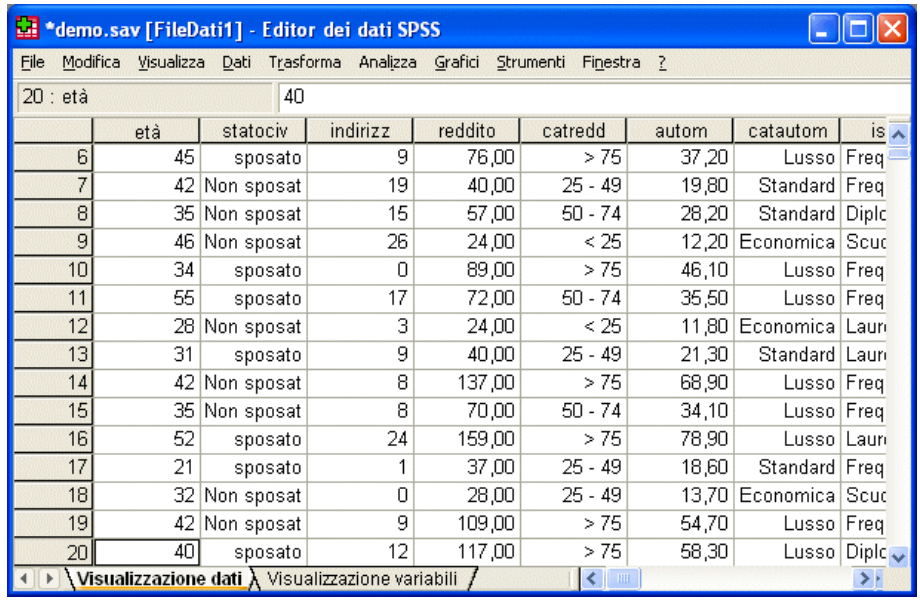
L'Editor dei dati offre due visualizzazioni dei dati:

- **Visualizzazione dati.** Visualizza i valori dei dati effettivi o le etichette dei valori definite.
- **Visualizzazione variabili.** Visualizza le informazioni di definizione delle variabili, incluse le etichette delle variabili e dei valori definite, il tipo di dati (ad esempio, stringa, data e numerico), il livello di misurazione (nominale, ordinale o scala) e i valori mancanti definiti dall'utente.

In entrambe le visualizzazioni, è possibile aggiungere, modificare ed eliminare le informazioni contenute nel file di dati.

Visualizzazione dati

Figura 5-1
Visualizzazione dati



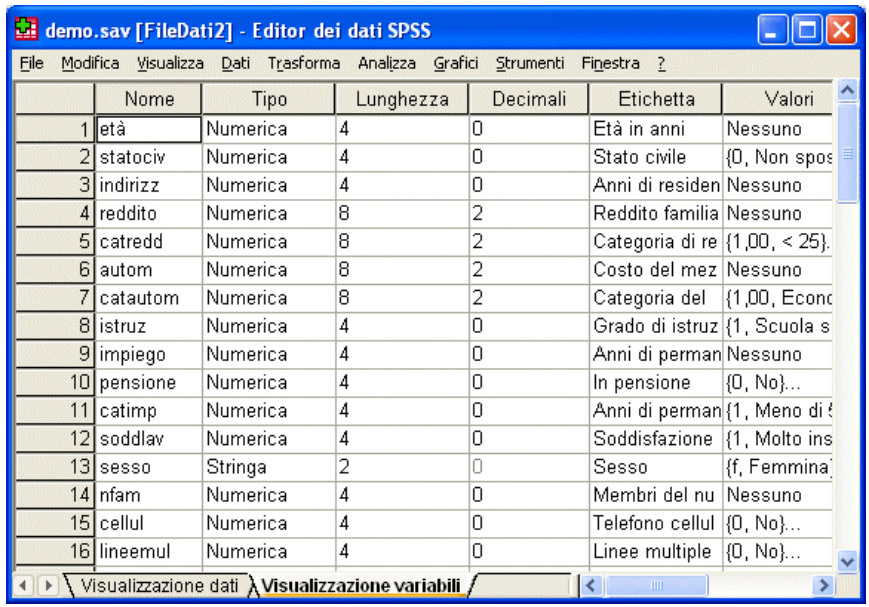
Molte funzioni della Visualizzazione dati sono analoghe a quelle disponibili nelle applicazioni per fogli di calcolo. Esistono, tuttavia, diverse importanti distinzioni:

- Le righe sono i casi. Ciascuna riga rappresenta un caso o un'osservazione. Ogni persona che risponde a un questionario, ad esempio, rappresenta un caso.
- Le colonne sono le variabili. Ciascuna colonna rappresenta una variabile o una caratteristica da misurare. Ogni voce del questionario, ad esempio, rappresenta una variabile.
- Le celle contengono valori. Ogni cella contiene un singolo valore di una variabile relativa a un caso. La cella è il punto in cui il caso e la variabile si intersecano. Le celle contengono solo valori. A differenza dei programmi per fogli di calcolo, le celle dell'Editor dei dati non possono contenere formule.
- Il file di dati è rettangolare. Le dimensioni del file di dati vengono determinate dal numero di casi e di variabili. È possibile inserire dati in qualsiasi cella. Se si inseriscono dati in una cella che non rientra nei limiti del file di dati definito, il rettangolo di dati verrà esteso in modo da includere le righe e/o le colonne

che si trovano tra la cella e i limiti del file. Entro i limiti del file di dati non sono presenti celle vuote. Per le variabili numeriche, le celle vuote vengono convertite al valore mancante di sistema. Per le variabili stringa, le celle vuote vengono considerate valori validi.

Visualizzazione variabili

Figura 5-2
Visualizzazione variabili


demo.sav [FileDati2] - Editor dei dati SPSS
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?
Visualizzazione dati
Visualizzazione variabili

	Nome	Tipo	Lunghezza	Decimali	Etichetta	Valori
1	età	Numerica	4	0	Età in anni	Nessuno
2	statociv	Numerica	4	0	Stato civile	{0, Non spos
3	indirizz	Numerica	4	0	Anni di residen	Nessuno
4	reddito	Numerica	8	2	Reddito familia	Nessuno
5	catredd	Numerica	8	2	Categoria di re	{1,00, < 25}.
6	autom	Numerica	8	2	Costo del mez	Nessuno
7	catautom	Numerica	8	2	Categoria del	{1,00, Econc
8	istruz	Numerica	4	0	Grado di istruz	{1, Scuola s
9	impiego	Numerica	4	0	Anni di perman	Nessuno
10	pensione	Numerica	4	0	In pensione	{0, No}...
11	catimp	Numerica	4	0	Anni di perman	{1, Meno di 5
12	soddlav	Numerica	4	0	Soddisfazione	{1, Molto ins
13	sesso	Stringa	2	0	Sesso	{f, Femmina}
14	nfam	Numerica	4	0	Membri del nu	Nessuno
15	cellul	Numerica	4	0	Telefono cellul	{0, No}...
16	lineemul	Numerica	4	0	Linee multiple	{0, No}...

La Visualizzazione variabili contiene le descrizioni degli attributi di ciascuna variabile del file di dati. Nella Visualizzazione variabili:

- Le righe sono le variabili.
- Le colonne sono gli attributi delle variabili.

È possibile aggiungere o eliminare variabili e modificare attributi delle variabili, tra cui:

- Nome di variabile
- Tipo di dati

- Numero di cifre o di caratteri
- Numero di posizioni decimali
- Etichette descrittive delle variabili e dei valori
- Valori mancanti definiti dall'utente
- Larghezza di colonna
- Livello di misurazione

Tutti questi attributi vengono salvati con il file di dati.

Oltre alla Visualizzazione variabili, esistono due ulteriori metodi per definire le proprietà delle variabili:

- La Copia proprietà dei dati guidata permette di usare un file di dati SPSS esterno o un altro file di dati disponibile nella sessione corrente come modello per la definizione delle proprietà dei file e delle variabli nel file di dati attivo. È possibile anche usare le variabile nel file di dati attivo come modelli per altre variabili dello stesso file. Copia proprietà dei dati è disponibile nel menu Dati della finestra dell'Editor dei dati.
- Definisci proprietà variabili (disponibile anche nel menu Dati della finestra dell'Editor dei dati) esamina i dati ed elenca tutti i valori univoci dei dati per ogni variabile selezionata, inoltre identifica i valori senza etichetta e fornisce una funzione per l'aggiunta automatica dell'etichetta. Questo risulta particolarmente utile per le variabili categoriali che utilizzano codici numerici per rappresentare le categorie, ad esempio 0 = *Maschio*, 1 = *Femmina*.

Per visualizzare e definire gli attributi delle variabili

▶ Attivare la finestra dell'Editor dei dati.

▶ Fare doppio clic su un nome di variabile nella parte superiore della colonna Visualizzazione dati oppure fare clic sulla scheda Visualizzazione variabili.

▶ Per definire una nuova variabile, immettere un nome di variabile in una qualsiasi riga vuota.

▶ Selezionare gli attributi da definire o modificare.

Nomi delle variabili

Per i nomi delle variabili sono valide le seguenti regole:

- Il nome deve iniziare con una lettera. Gli altri caratteri possono essere lettere, cifre, punti o i simboli @, #, _ , o $.
- I nomi delle variabili non possono terminare con un punto.
- Evitare nomi di variabili che terminano con un carattere di sottolineatura (per evitare conflitti con le variabili create automaticamente da alcune procedure).
- Il nome non può avere lunghezza superiore a 64 byte. 64 byte in genere corrispondono a 64 caratteri nelle lingue a un byte (ad esempio, inglese, francese, tedesco, spagnolo, italiano, ebraico, russo, greco, arabo, tailandese) e a 32 caratteri nelle lingue a due byte (giapponese, cinese, coreano).
- Non è possibile utilizzare spazi e caratteri speciali, ad esempio !, ?, ’e *.
- Ogni variabile deve essere univoca. La duplicazione non è consentita.
- Le parole chiave standard non possono essere utilizzate come nomi di variabili. Le parole chiave standard sono: `ALL`, `AND`, `BY`, `EQ`, `GE`, `GT`, `LE`, `LT`, `NE`, `NOT`, `OR`, `TO`, `WITH`.
- È possibile definire i nomi delle variabili con un misto di caratteri maiuscoli e minuscoli e la distinzione viene mantenuta per motivi di visualizzazione.
- Quando è necessario impostare il ritorno a capo su più righe dell’output per i nomi di variabili troppo lunghi, SPSS consente di interrompere le righe in corrispondenza dei caratteri di sottolineatura, dei punti e modifica i caratteri minuscoli in maiuscolo.

Livello di misurazione delle variabili

È possibile specificare il livello di misurazione in forma di scala (dati numerici su una scala di intervallo o di rapporto), come ordinale o nominale. I dati nominali e ordinali possono essere rappresentati da stringhe (alfanumeriche) o essere numerici.

- **Nominale.** Una variabile può essere considerata nominale quando i suoi valori rappresentano categorie prive di ordinamento intrinseco; per esempio l'ufficio di una società, la regione, il codice postale, la religione.

- **Ordinale.** Una variabile può essere considerata ordinale quando i suoi valori rappresentano categorie con qualche ordinamento intrinseco; per esempio i gradi di soddisfazione per un servizio, da molto insoddisfatto a molto soddisfatto, o punteggi di preferenza.
- **Scala.** Una variabile può essere considerata di scala quando i suoi valori rappresentano categorie ordinate con una metrica significativa, tale che siano appropriati confronti fra le distanze dei suoi valori. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

Nota: per le variabili stringa ordinali si suppone che l'ordine alfabetico dei valori stringa rifletta l'esatto ordine delle categorie. Ad esempio, per una variabile stringa con i valori *basso*, *medio*, *alto*, l'ordine delle categorie viene interpretato come *alto*, *basso*, *medio*, ma questo non è l'ordine corretto. In generale, per rappresentare i dati ordinali, è più sicuro utilizzare i codici numerici.

Per le nuove variabili create durante una sessione, il livello di misurazione predefinito assegnato ai dati letti dai file esterni e dai file di dati SPSS di versioni precedenti alla 8 è basato sulle seguenti regole:

- Le variabili numeriche con meno di 24 valori univoci e le variabili stringa sono impostate su Nomimale.
- Le variabili numeriche con 24 o più valori univochi sono impostate su Scala.

Il valore di riferimento scala/nominale per le variabili scala/nominale può essere modificato nella finestra di dialogo Opzioni. Per ulteriori informazioni, vedere "Opzioni dei grafici interattivi" in Capitolo 44 a pag. 641.

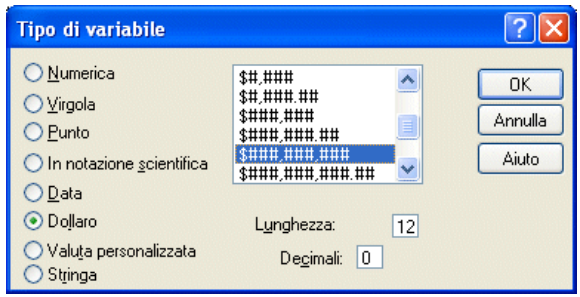
La finestra di dialogo Definisci proprietà variabili nel menu Dati permette di assegnare più agevolmente il livello di misurazione corretto. Per ulteriori informazioni, vedere "Assegnazione del livello di misurazione" in Capitolo 7 a pag. 129.

Tipo di variabile

Tipo di variabile consente di specificare il tipo di dati per ciascuna variabile. Per impostazione predefinita, si presume che tutte le nuove variabili siano numeriche. È possibile utilizzare Tipo di variabile per modificare il tipo di dati. Il contenuto della finestra di dialogo Tipo di variabile varia in base al tipo di dati selezionato. Per alcuni

tipi di dati, sono disponibili caselle di testo per immettere la larghezza e il numero di decimali, mentre per altri è sufficiente selezionare un formato da un elenco di esempi.

Figura 5-3
Finestra di dialogo Tipo di variabile



I tipi di dati disponibili sono:

Numerica. Una variabile che contiene solo dati numerici. visualizzati nel formato numerico standard. L'Editor dei dati accetta valori numerici nel formato standard oppure in notazione scientifica.

Virgola. Una variabile numerica i cui valori sono delimitati con una virgola ogni tre posizioni e con il punto come delimitatore dei decimali. Per questo tipo di variabili l'Editor dei dati accetta valori numerici con o senza virgola oppure valori in notazione scientifica. I valori non possono contenere virgole a destra dell'indicatore decimale.

Punto. Una variabile numerica i cui valori sono delimitati con un punto ogni tre posizioni e con la virgola come delimitatore dei decimali. L'editor dei dati accetta valori numerici per variabili punto con o senza punti oppure li accetta in notazione scientifica. I valori non possono contenere punti a destra dell'indicatore decimale.

In notazione scientifica. Una variabile numerica i cui valori sono visualizzati con una E e un esponente alla decima potenza. L'editor dei dati accetta valori numerici per tali variabili con o senza esponente. L'esponente può essere preceduto da E o D con un simbolo opzionale o dal solo simbolo, ad esempio 123, 1.23E2, 1.23D2, 1.23E+2 o anche 1.23+2.

Data. Una variabile numerica i cui valori sono visualizzati in uno dei formati calendario-data o orologio-tempo. Selezionare un formato dall'elenco. I delimitatori di data accettati sono barre, trattini, punti, virgole o spazi vuoti. L'intervallo di secoli per i valori di anni a due cifre viene determinato mediante le impostazioni delle opzioni (menu Modifica, Opzioni, schedaDati).

Dollaro. Una variabile numerica visualizzata con un simbolo dollaro iniziale ($), con virgole di delimitazione per ogni tre spazi e un punto come delimitatore decimale. È possibile immettere i valori dati con o senza il simbolo dollaro iniziale.

Valuta personalizzata. Una variabile numerica i cui valori sono visualizzati in uno dei formati di valuta personalizzati definiti nella scheda Valuta della finestra di dialogo Opzioni. I caratteri di valuta personalizzati definiti non possono essere utilizzati nell'inserimento dei dati ma vengono visualizzati nell'Editor dei dati.

Stringa. I valori di una variabile stringa non sono numerici e non vengono quindi utilizzati per i calcoli. Queste variabili possono contenere qualsiasi carattere e non devono superare la lunghezza definita. È attiva la distinzione tra maiuscole e minuscole. Questo tipo di variabili è chiamato anche variabili alfanumeriche.

Per definire il tipo di variabile

- ▶ Fare clic sul pulsante della cella *Tipo* della variabile che si intende definire.
- ▶ Selezionare il tipo di dati desiderato nella finestra di dialogo Tipo di variabile.
- ▶ Fare clic su OK.

Formati di input e di visualizzazione

A seconda del formato, i valori nella Visualizzazione dati possono differire dal valore effettivo inserito e memorizzato internamente. Di seguito vengono riportate alcune indicazioni generali:

- Per i formati numerico, virgola e punto, è possibile specificare valori con qualsiasi numero di cifre decimali (fino a 16). Il valore verrà memorizzato per intero. Nella Visualizzazione dati verrà visualizzato solo il numero specificato di cifre decimali. I valori con cifre decimali in eccesso verranno arrotondati. In qualsiasi calcolo verrà tuttavia utilizzato il valore completo.
- Per le variabili stringa, a tutti i valori viene aggiunto il numero di spazi necessario per raggiungere la larghezza massima. Per una variabile stringa con una larghezza massima pari a tre, il valore *No* viene salvato internamente come `'No '` ed è equivalente a `' No'`.
- Per i formati della data, è possibile utilizzare barre, trattini, spazi, virgole o punti come delimitatori tra i valori relativi a giorno, mese e anno, nonché aggiungere numeri, abbreviazioni di tre lettere o nomi completi per i valori relativi ai mesi.

Le date espresse nel formato generico gg-mmm-aa vengono visualizzate con il trattino come delimitatore e l'abbreviazione di tre lettere del mese. Le date espresse nel formato generico gg/mm/aa e mm/gg/aa vengono visualizzate con la barra come delimitatore e i numeri corrispondenti ai mesi. Le date vengono memorizzate internamente come numero di secondi trascorsi dal 14 ottobre 1582. L'intervallo di secoli per gli anni a due cifre viene determinato mediante le impostazioni delle opzioni (menu Modifica, Opzioni, scheda Dati).

- Per i formati dell'orario, è possibile utilizzare i due punti, il punto o lo spazio come delimitatore tra ore, minuti e secondi. Gli orari vengono visualizzati con i due punti come delimitatore. Internamente le date vengono memorizzate come il numero di secondi trascorsi dal 14 ottobre 1582.

Etichette delle variabili

È possibile assegnare etichette variabili descrittive della lunghezza massima di 256 caratteri (o 128 caratteri nelle lingue a due byte). Le etichette variabili possono contenere spazi e caratteri riservati non consentiti nei nomi delle variabili.

Per specificare le etichette di variabile

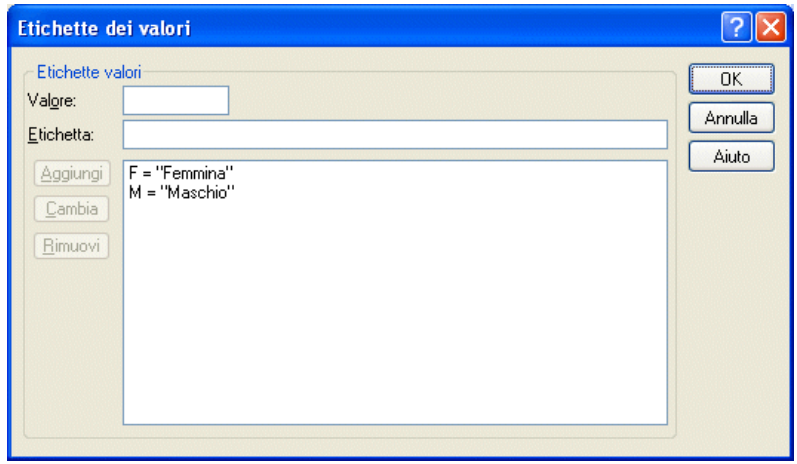
- Attivare la finestra dell'Editor dei dati.
- Fare doppio clic su un nome di variabile nella parte superiore della colonna Visualizzazione dati oppure fare clic sulla scheda Visualizzazione variabili.
- Immettere l'etichetta di variabile descrittiva nella cella *Etichetta*.

Etichette dei valori

È possibile assegnare etichette descrittive a ciascun valore di una variabile. Questo processo risulta particolarmente utile se il file di dati utilizza codici numerici per rappresentare categorie non numeriche, ad esempio i codici 1 e 2 per *maschio* e *femmina*.

- Le etichette dei valori possono essere formate al massimo da 120 byte.
- Le etichette dei valori non sono disponibili per le variabili stringa lunga (variabili stringa formate da più di 8 caratteri).

Figura 5-4
Finestra di dialogo Etichette dei valori.



Per specificare le etichette dei valori

- Fare clic sul pulsante della cella *Valori* della variabile che si intende definire.
- Per ciascun valore, specificare il valore e l'etichetta corrispondente.
- Fare clic su Aggiungi per inserire l'etichetta del valore.
- Fare clic su OK.

Inserimento di interruzioni di riga nelle etichette

Per le etichette di variabili e dei valori, viene impostato il ritorno a capo automatico su più righe nelle tabelle pivot e nei grafici se la larghezza della cella o dell'area non consente di visualizzare l'intera etichetta su un'unica linea. Per spostare il ritorno a capo in un punto diverso dell'etichetta, è possibile modificare i risultati e inserire manualmente le interruzioni di riga. È inoltre possibile creare etichette di variabili e dei valori in cui il ritorno a capo si verificherà *sempre* in determinati punti con conseguente visualizzazione su più righe:

- Per le etichette delle variabili, selezionare la cella Etichetta per la variabile nella Visualizzazione variabili dell'Editor dei dati.

- Per le etichette dei valori, selezionare la cella Valori per la variabile nella Visualizzazione variabili dell'Editor dei dati, fare clic sul pulsante visualizzato nella cella e selezionare l'etichetta che si desidera modificare nella finestra di dialogo Etichette dei valori.

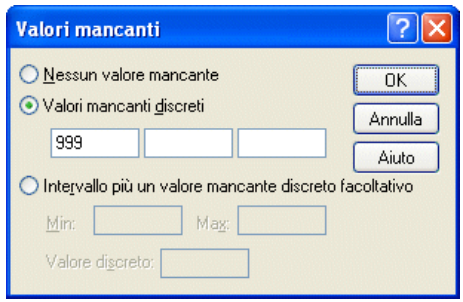
- Digitare \n all'interno dell'etichetta nel punto in cui inserire l'interruzione di riga.

  Il simbolo \n non viene visualizzato nelle tabelle pivot o nei grafici, ma viene considerato un carattere indicante l'interruzione di riga.

Valori mancanti

Definisci valori mancanti consente di determinare i valori dei dati specificati come **valori mancanti definiti dall'utente**. Può essere necessario, ad esempio, distinguere tra dati che mancano perché la persona ha rifiutato di rispondere e dati che mancano perché la domanda non riguardava la persona in questione. I dati specificati come valori mancanti definiti dall'utente vengono evidenziati in modo da poter essere elaborati diversamente ed esclusi dalla maggior parte dei calcoli.

Figura 5-5
Finestra di dialogo Valori mancanti



- È possibile inserire al massimo tre valori mancanti discreti (singoli), un intervallo di valori mancanti oppure un intervallo più un valore discreto.
- È possibile specificare intervalli solo per le variabili numeriche.
- Non è possibile definire valori mancanti per le variabili stringa lunga (variabili stringa formate da più di 8 caratteri).

Valori mancanti nelle variabili stringa. Tutti i valori delle stringhe, inclusi i valori nulli o vuoti, sono considerati validi a meno che non siano esplicitamente definiti mancanti. Per definire come valori mancanti i valori nulli o vuoti di una variabile stringa, immettere uno spazio singolo in uno dei campi per Valori mancanti discreti.

Per definire i valori mancanti

- Fare clic sul pulsante della cella *Mancante* per la variabile che si intende definire.
- Inserire i valori o l'intervallo di valori che rappresentano i dati mancanti.

  Tutti i valori delle stringhe, inclusi i valori nulli o vuoti, sono considerati validi a meno che non siano esplicitamente definiti mancanti. Per definire come valori mancanti i valori nulli o vuoti di una variabile stringa, immettere uno spazio singolo in uno dei campi per Valori mancanti discreti.

Larghezza colonna

È possibile specificare un numero di caratteri per la larghezza della colonna. Per modificare la larghezza delle colonne nella Visualizzazione dati, fare clic e trascinare i bordi delle colonne.

I formati delle colonne influiscono solo sulla visualizzazione dei valori nell'Editor di dati. La modifica della larghezza delle colonne non comporta la modifica della larghezza definita di una variabile. Se le misure della larghezza definita ed effettiva di un valore sono maggiori di quella della colonna, nella Visualizzazione dati verranno visualizzati degli asterischi (*).

Allineamento

Allineamento consente di controllare la visualizzazione dei valori dei dati e/o delle etichette dei valori nella Visualizzazione dati. L'allineamento predefinito è a destra per le variabili numeriche e a sinistra per le variabili stringa. L'impostazione influisce esclusivamente sulla Visualizzazione dati.

Applicazione degli attributi di definizione di una variabile ad altre variabili

Dopo avere specificato gli attributi di definizione di una variabile, è possibile copiare uno o più attributi e applicarli a una o più variabili.

Per applicare gli attributi di definizione di una variabile, è possibile utilizzare le operazioni di copia e incolla di base. È possibile:

- Copiare un singolo attributo (ad esempio, le etichette dei valori) e incollarlo nella cella del medesimo attributo di una o più variabili.
- Copiare tutti gli attributi di una variabile e incollarli in una o più variabili diverse.
- Creare più variabili nuove con tutti gli attributi di una variabile copiata.

Applicazione degli attributi di definizione di una variabile ad altre variabili

Per applicare dei singoli attributi di una determinata variabile:

▶ Nella Visualizzazione variabili, selezionare la cella dell'attributo da applicare alle altre variabili.

▶ Dai menu, scegliere:

Modifica
  Copia

▶ Selezionare la cella a cui applicare l'attributo. È possibile selezionare più variabili di destinazione.

▶ Dai menu, scegliere:

Modifica
  Incolla

Se si incolla l'attributo in righe vuote, vengono create nuove variabili con i valori predefiniti per tutti gli attributi fatta eccezione per l'attributo selezionato.

Per applicare tutti gli attributi di una determinata variabile:

▶ Nella Visualizzazione variabili, selezionare il numero di riga della variabile con gli attributi da utilizzare. Verrà evidenziata l'intera riga.

▶ Dai menu, scegliere:

Modifica
  Copia

▶ Selezionare i numeri di riga delle variabili a cui applicare gli attributi. È possibile selezionare più variabili di destinazione.

▶ Dai menu, scegliere:

Modifica
  Incolla

Creazione di più variabili nuove con gli stessi attributi

▶ Nella Visualizzazione variabili, fare clic sul numero di riga della variabile con gli attributi da utilizzare per la nuova variabile. Verrà evidenziata l'intera riga.

▶ Dai menu, scegliere:

Modifica
  Copia

▶ Fare clic sul numero di riga vuota sotto l'ultima variabile definita nel file di dati.

▶ Dai menu, scegliere:

Modifica
  Incolla variabili...

▶ Nella finestra di dialogo Incolla variabili, immettere il numero di variabili da creare.

▶ Inserire un prefisso e un numero iniziali per le nuove variabili.

▶ Fare clic su OK.

I nomi delle nuove variabili saranno composti dal prefisso specificato a cui viene aggiunto un numero sequenziale che inizia con il numero specificato.

Inserimento di dati

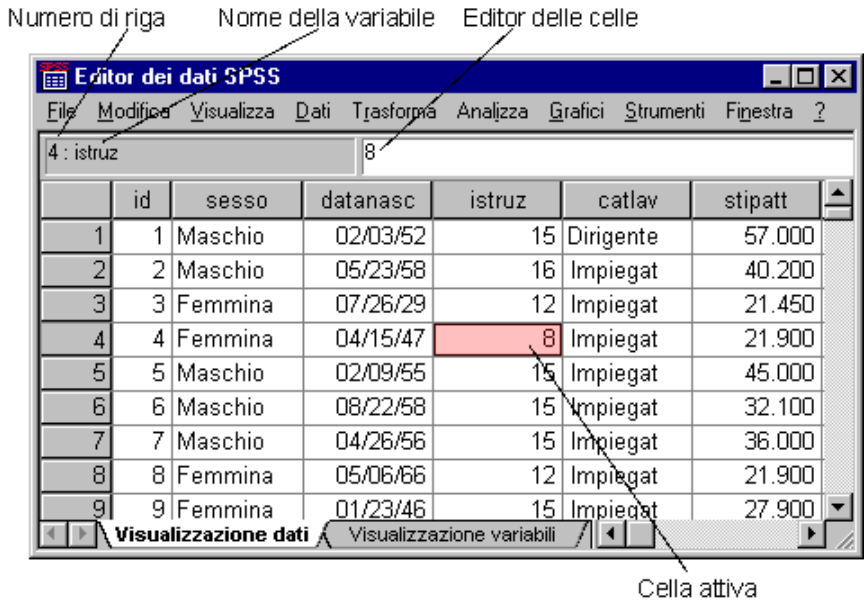
Nella Visualizzazione dati è possibile immettere i dati direttamente nell'Editor di dati. I dati possono essere inseriti in qualsiasi ordine. I dati possono essere inseriti per caso o per variabile nelle aree selezionate o nelle singole celle.

- La cella attiva viene evidenziata.
- Il nome della variabile e il numero di riga della cella attiva verranno visualizzati nell'angolo in alto a sinistra dell'Editor di dati.
- Quando si seleziona una cella e si inserisce un valore, questo verrà visualizzato nell'editor delle celle nella parte superiore dell'Editor dei dati.

- I valori dei dati verranno registrati solo quando si preme Invio o si seleziona un'altra cella.
- Per inserire dati diversi da semplici dati numerici, è innanzitutto necessario definire il tipo di variabile.

Se si inserisce un valore in una colonna vuota, verrà creata automaticamente una nuova variabile a cui verrà assegnato un nome.

Figura 5-6
File di dati attivo nella Visualizzazione dei dati


Numero di riga
Nome della variabile
Editor delle celle
Editor dei dati SPSS
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?
4 : istruz
8
id sesso datanasc istruz catlav stipatt
1 1 Maschio 02/03/52 15 Dirigente 57.000
2 2 Maschio 05/23/58 16 Impiegat 40.200
3 3 Femmina 07/26/29 12 Impiegat 21.450
4 4 Femmina 04/15/47 8 Impiegat 21.900
5 5 Maschio 02/09/55 15 Impiegat 45.000
6 6 Maschio 08/22/58 15 Impiegat 32.100
7 7 Maschio 04/26/56 15 Impiegat 36.000
8 8 Femmina 05/06/66 12 Impiegat 21.900
9 9 Femmina 01/23/46 15 Impiegat 27.900
Visualizzazione dati
Visualizzazione variabili
Cella attiva

Per inserire dati numerici

- ▶ Selezionare una cella nella Visualizzazione dati.
- ▶ Inserire il valore. Il valore viene visualizzato nell'editor celle nella parte superiore dell'Editor di dati.
- ▶ Per registrare il valore, premere Invio oppure selezionare un'altra cella.

Per inserire dati non numerici

- Fare doppio clic su un nome di variabile nella parte superiore della colonna Visualizzazione dati oppure fare clic sulla scheda Visualizzazione variabili.
- Fare clic sul pulsante nella cella *Tipo* per la variabile.
- Selezionare il tipo di dati desiderato nella finestra di dialogo Tipo di variabile.
- Fare clic su OK.
- Fare doppio clic sul numero di riga oppure sulla scheda Visualizzazione dati.
- Inserire i dati nella colonna della nuova variabile definita.

Per utilizzare etichette dei valori per l'inserimento dei dati

- Se nella Visualizzazione dati non sono visualizzate le etichette dei valori, scegliere:
  Visualizza
  Etichette dei valori
- Fare clic sulla cella in cui immettere il valore.
- Selezionare un'etichetta dall'elenco a discesa.

  Il valore verrà inserito e l'etichetta verrà visualizzata nella cella.

  Nota: Questo processo funziona solo se per la variabile sono state definite etichette di valori.

Limitazioni relative ai valori

Il tipo di valore che può essere inserito nella cella nella Visualizzazione dati viene determinato in base al tipo di variabile e alla larghezza definiti.

- Se si digita un carattere non consentito dal tipo di variabile definito, nell'Editor dei dati verrà emesso un segnale acustico e il carattere non verrà inserito.
- Nelle variabili stringa non è consentito inserire un numero di caratteri superiore a quello definito.
- Nelle variabili numeriche è possibile inserire valori interi maggiori della larghezza definita, ma nell'Editor dei dati verranno visualizzati asterischi o una notazione scientifica nella cella ad indicare che il valore supera la larghezza definita. Per

visualizzare il valore nella cella, è necessario modificare la larghezza definita per la variabile. (*Nota*: La modifica della larghezza della colonna non comporta la modifica della lunghezza della variabile.

Modifica dei dati nella Visualizzazione dati

Con l'Editor dei dati è possibile modificare i valori dei dati nella Visualizzazione dati in diversi modi. È possibile:

- Modificare i valori.
- Tagliare, copiare e incollare valori.
- Aggiungere ed eliminare casi.
- Aggiungere ed eliminare variabili.
- Modificare l'ordine delle variabili.

Sostituzione o modifica di valori

Per eliminare il vecchio valore e inserire un nuovo valore:

- Nella Visualizzazione dati, fare doppio clic sulla cella. Il valore della cella verrà visualizzato nell'editor delle celle.
- Modificare il valore direttamente nella cella o nell'editor delle celle.
- Premere Invio (oppure passare a un'altra cella) per registrare il nuovo valore.

Tagliare, copiare e incollare valori

Nell'Editor dei dati è possibile tagliare, copiare e incollare i valori delle singole celle oppure gruppi di valori. È possibile:

- Spostare o copiare il valore di una cella in un'altra cella.
- Spostare o copiare il valore di una cella in un gruppo di celle.
- Spostare o copiare i valori di un singolo caso (riga) in più casi.
- Spostare o copiare i valori di una singola variabile (colonna) in più variabili.
- Spostare o copiare i valori di un gruppo di celle in un altro gruppo di celle.

Conversione di dati per i valori incollati nella Visualizzazione dati

Se i tipi di variabile definiti delle celle di origine non coincidono con quelli delle celle di destinazione, verrà eseguito un tentativo di conversione del valore. Se la conversione non è possibile, nella cella di destinazione verrà inserito il valore mancante di sistema.

Conversione di stringhe in valori numerici o date. I formati numerici (ad esempio numerico, dollaro, punto o virgola) e della data verranno convertiti in stringhe se incollati nella cella di una variabile stringa. Il valore stringa corrisponde al valore numerico visualizzato nella cella. Per una variabile in formato dollaro, il segno di dollaro visualizzato diventa parte del valore stringa. I valori di larghezza superiore a quella definita per la variabile stringa verranno troncati.

Conversione di stringhe in valori numerici o date. Le stringhe contenenti caratteri consentiti per il formato numerico o della data della cella di destinazione verranno convertiti nel corrispondente valore numerico o data. Il valore stringa 25/12/91, ad esempio, viene convertito in una data valida se il tipo di formato della cella di destinazione è uno dei formati giorno-mese-anno; se invece quest'ultimo è uno dei formati mese-giorno-anno, il valore verrà convertito in valore mancante di sistema.

Conversione di date in valori numerici. I valori relativi alla data e all'ora vengono convertiti in un numero di secondi se la cella di destinazione corrisponde a un formato numerico (ad esempio, numerico, dollaro, punto o virgola). Poiché le date vengono memorizzate internamente come il numero di secondi trascorsi dal 14 ottobre 1582, la conversione delle date in valori numerici può produrre risultati composti da un numero considerevole di cifre. La data 29/10/91, ad esempio, viene convertita nel valore numerico 12.908.073.600.

Conversione di valori numerici in date o ore. I valori numerici vengono convertiti in date o orari se il valore rappresenta un numero di secondi che può produrre una data o un orario valido. Per quanto riguarda le date, i valori numerici inferiori a 86.400 verranno convertiti nel valore mancante di sistema.

Inserimento di nuovi casi

Se si immettono dati in una cella o in una riga vuota verrà creato automaticamente un nuovo caso. Verrà inserito automaticamente il valore mancante di sistema per tutte le altre variabili del caso specifico. Se tra il nuovo caso e i casi esistenti sono

interposti spazi vuoti, anche le righe vuote diventeranno nuovi casi con il valore mancante di sistema specificato per tutte le variabili. È inoltre possibile inserire nuovi casi tra casi esistenti.

Per inserire un nuovo caso tra casi esistenti

- Nella Visualizzazione dati, selezionare una cella qualsiasi nel caso (riga) sottostante la posizione in cui si desidera inserire un nuovo caso.
- Dai menu, scegliere:

  Modifica
  Inserisci casi

  Verrà inserita una nuova riga per il caso e in tutte le variabili verrà immesso il valore mancante di sistema.

Inserimento di nuove variabili

L'immissione di dati in una colonna vuota nella Visualizzazione dati o in una riga vuota nella Visualizzazione variabili comporta la creazione automatica di una nuova variabile con un nome predefinito (il prefisso *var* e un numero progressivo) e di un tipo di formato dei dati predefinito (numerico). Per tutti i casi della nuova variabile verrà inserito automaticamente il valore mancante di sistema. Se nella Visualizzazione dati o nella Visualizzazione variabili vi sono rispettivamente delle colonne o delle righe vuote comprese tra la nuova variabile e quelle esistenti, anche queste divengono nuove variabili con il valore mancante di sistema per tutti i casi. È inoltre possibile inserire una nuova variabile tra le variabili esistenti.

Per inserire una nuova variabile tra le variabili esistenti

- Selezionare qualsiasi cella della variabile a destra (Visualizzazione dati) o sotto (Visualizzazione variabili) la posizione in cui inserire la nuova variabile.
- Dai menu, scegliere:

  Modifica
  Inserisci variabile

  Verrà inserita una nuova variabile con il valore mancante di sistema per tutti i casi.

Per spostare le variabili

- Per selezionare la variabile, fare clic sul nome della variabile nella Visualizzazione dati o sul numero di riga della variabile nella Visualizzazione variabili.
- Trascinare la variabile nella nuova posizione.
- Per inserire la variabile tra due variabili esistenti: nella Visualizzazione dati, inserire la variabile nella colonna variabili a destra del punto in cui si desidera inserirla oppure nella Visualizzazione variabili inserirla nella riga variabili sotto al punto in cui si desidera inserirla.

Per modificare il tipo di dati

È possibile modificare il tipo di dati di una variabile in qualunque momento tramite la finestra di dialogo Tipo di variabile nella Visualizzazione variabili. L'Editor di dati tenta di convertire i valori esistenti nel nuovo tipo. Se la conversione non è possibile, verrà assegnato il valore mancante di sistema. Le regole di conversione sono analoghe a quelle utilizzate per incollare valori in una variabile con un formato diverso. Se la modifica del formato dei dati comporta la perdita delle specifiche dei valori mancanti o delle etichette dei valori, verrà visualizzata una finestra di messaggio in cui viene richiesto se procedere alla modifica o annullarla.

Vai al caso

La finestra di dialogo Vai al caso permette di andare al numero di caso specificato (riga) nell'Editor di dati.

Figura 5-7
Finestra di dialogo Vai al caso.

Per cercare un caso nell'Editor dei dati

- Attivare la finestra dell'Editor dei dati.

- Dai menu, scegliere:

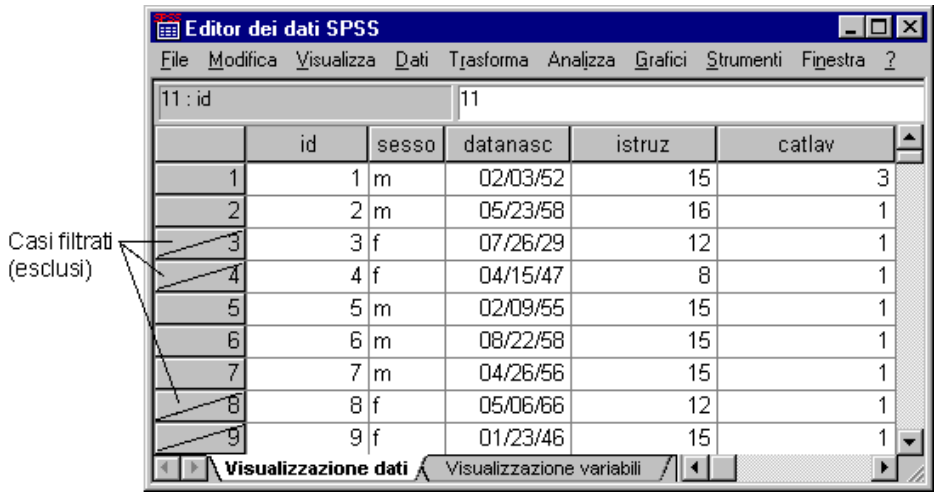
  Dati
  Vai al caso...

- Nella finestra di dialogo Vai al caso, immettere il numero di riga dell'Editor di dati per il caso.

Criteri di selezione dei casi nell'Editor dei dati

Se si è selezionato un sottoinsieme di casi senza eliminare i casi non selezionati, nell'Editor dei dati questi ultimi verranno contrassegnati da una linea diagonale (barra) sopra il numero di riga.

Figura 5-8
Casi filtrati nell'Editor dei dati


Editor dei dati SPSS
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?
11 : id
11
id sesso datanasc istruz catlav
1 1 m 02/03/52 15 3
2 2 m 05/23/58 16 1
3 3 f 07/26/29 12 1
4 4 f 04/15/47 8 1
5 5 m 02/09/55 15 1
6 6 m 08/22/58 15 1
7 7 m 04/26/56 15 1
8 8 f 05/06/66 12 1
9 9 f 01/23/46 15 1
Visualizzazione dati
Visualizzazione variabili
Casi filtrati
(esclusi)

Editor dei dati: Opzioni di visualizzazione

Nel menu Visualizza sono disponibili diverse opzioni di visualizzazione per l'Editor dei dati:

Caratteri. Questa opzione consente di controllare le caratteristiche dei caratteri per la visualizzazione dei dati.

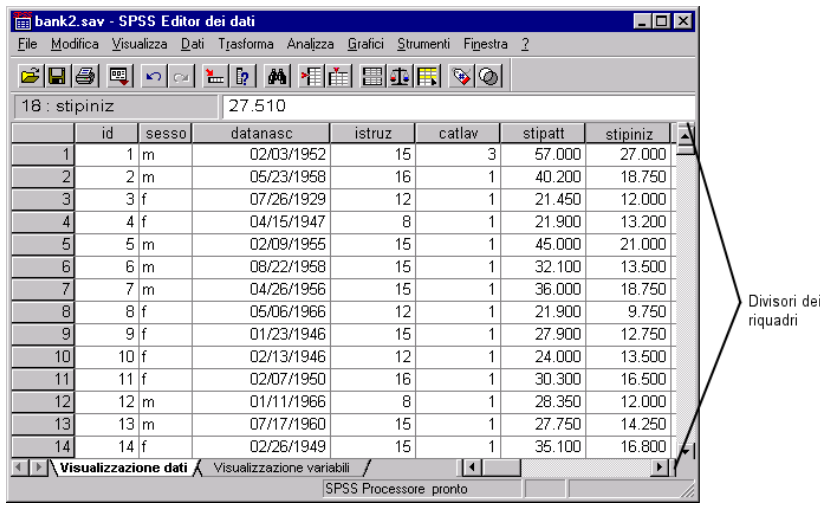
Griglia. Questa opzione attiva o disattiva la visualizzazione delle linee della griglia.

Etichette dei valori. Questa opzione consente di visualizzare alternativamente i dati effettivi e le etichette descrittive dei valori definite dall'utente. Ciò è possibile solo nella Visualizzazione dati.

Uso delle viste multiple

In Vista dati è possibile creare viste multiple (riquadri) utilizzando i divisori situati sopra la barra di scorrimento orizzontale e a destra della barra di scorrimento verticale.

Figura 5-9
Divisori del riquadro Vista dati



È possibile anche utilizzare il menu Finestra per inserire e rimuovere i divisori dei riquadri. Per inserire i divisori:

▶ Nei menu della Visualizzazione dati scegliere:

Finestra
  Dividi

I divisori vengono inseriti sopra e a sinistra della cella selezionata.

- Se si seleziona la cella in alto a sinistra, i divisori inseriti dividono la vista corrente a circa metà, sia orizzontalmente che verticalmente.

- Se si seleziona una cella diversa dalla prima cella della prima colonna, viene inserito un divisore di riquadri orizzontale sopra la cella selezionata.
- Se si seleziona una cella diversa dalla prima cella della prima riga, viene inserito un divisore di riquadri verticale a sinistra della cella selezionata.

Stampa dall'Editor dei dati

I file dati vengono stampati esattamente come vengono visualizzati.

- Vengono stampati i dati presenti nella vista correntemente visualizzata. Nella Visualizzazione dati, vengono stampati i dati. Nella Visualizzazioni variabili, vengono stampate le informazioni di definizione dei dati.
- Le griglie vengono stampate se visualizzate nella vista selezionata.
- Le etichette dei valori vengono stampate nella Visualizzazione dati se sono visualizzate. In caso contrario, vengono stampati i valori dei dati effettivi.

Utilizzare il menu Visualizza nell'Editor dei dati per visualizzare o nascondere la griglia e impostare la visualizzazione dei valori o delle etichette dei valori.

Per stampare il contenuto dell'Editor dei dati

▶ Attivare la finestra dell'Editor dei dati.

▶ Selezionare la scheda relativa alla visualizzazione da stampare.

▶ Dai menu, scegliere:

File
  Stampa...

Capitolo
6

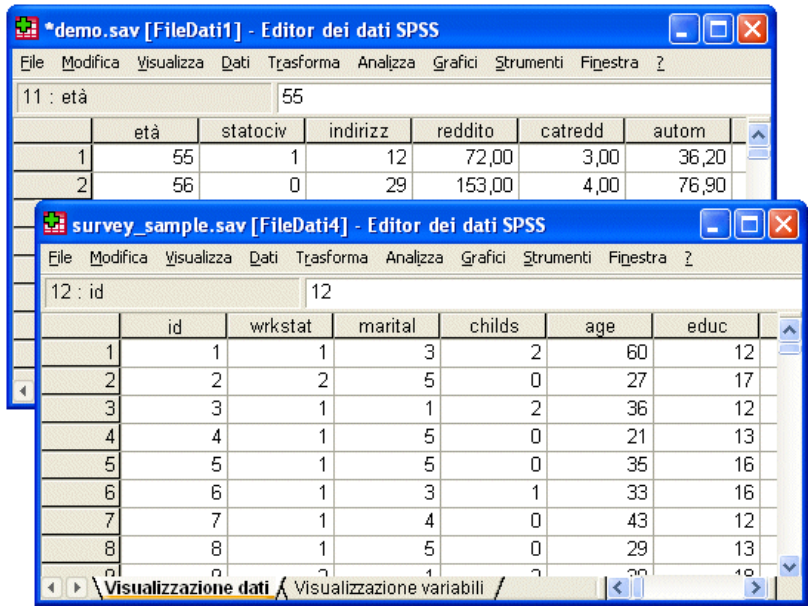
Utilizzo di più sorgenti dati

A partire dalla versione SPSS 14.0, SPSS consente di tenere aperte più sorgenti dati contemporaneamente. Ciò permette agevolmente di:

- Passare da una sorgente dati a un'altra.
- Confrontare i contenuti di più sorgenti dati.
- Copiare e incollare i dati tra le sorgenti dati.
- Creare più sottoinsiemi di casi e/o di variabili per l'analisi.
- Unire più sorgenti di dati di formati dati diversi (ad esempio fogli di calcolo, database e dati di testo) senza dover salvare ciascuna sorgente di dati in formato SPSS.

Gestione di più sorgenti di dati

Figura 6-1
Due sorgenti di dati aperte contemporaneamente

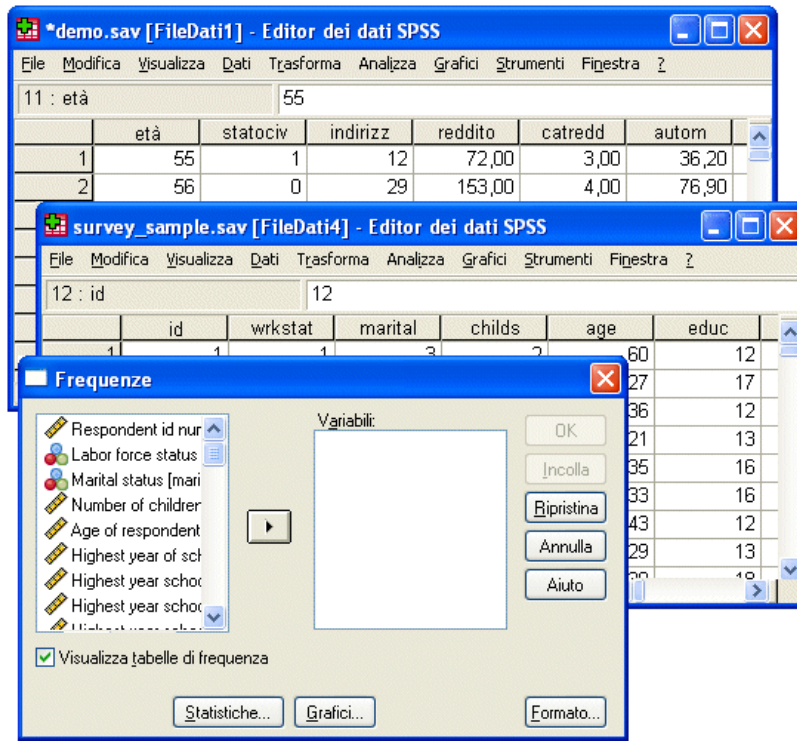

*demo.sav [FileDati1] - Editor dei dati SPSS
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?
11 : età
55
età statociv indirizz reddito catredd autom
1 55 1 12 72,00 3,00 36,20
2 56 0 29 153,00 4,00 76,90
survey_sample.sav [FileDati4] - Editor dei dati SPSS
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?
12 : id
12
id wrkstat marital childs age educ
1 1 1 3 2 60 12
2 2 2 5 0 27 17
3 3 1 1 2 36 12
4 4 1 5 0 21 13
5 5 1 5 0 35 16
6 6 1 3 1 33 16
7 7 1 4 0 43 12
8 8 1 5 0 29 13
Visualizzazione dati
Visualizzazione variabili

Le sorgenti dati aperte vengono visualizzate in una nuova finestra dell'Editor dei dati.

- Tutte le sorgenti di dati già aperte rimangono tali e possono essere usate.
- La sorgente di dati aperte diventa automaticamente il **file di dati attivo**.
- Per modificare il file di dati attivo, è sufficiente fare clic in qualsiasi punto della finestra dell'Editor di dati della sorgente di dati da usare oppure selezionare la finestra Editor di dati per la sorgente di dati dal menu Finestra.

- Solo le variabili del file di dati attivo possono essere usate per l'analisi.

Figura 6-2
Elenco di variabili contenente le variabili del file di dati attivo



- Il file di dati attivo non può essere modificato se è aperta una finestra di dialogo che ha accesso ai dati (ciò include anche le finestre di dialogo che visualizzano gli elenchi di variabili).
- È sempre necessario aprire almeno una finestra dell'Editor di dati per sessione. Quando si chiude l'ultima finestra dell'Editor di dati aperta, SPSS la chiude automaticamente chiedendo all'utente se desidera salvare le modifiche.

Nota: se si utilizza la sintassi del comando per aprire le sorgenti di dati (ad esempio`GET FILE`, `GET DATA`), è necessario nominare esplicitamente ciascun file di dati per poter tenere contemporaneamente aperte più sorgenti di dati.

Copia e inserimento di informazioni in più file di dati

È possibile copiare gli attributi delle definizioni dei dati e delle variabili da un insieme di dati a un altro eseguendo le stesse operazioni utilizzabili all'interno dei singoli file di dati.

- Se si copiano e si incollano le celle dati selezionati in Visualizzazione dei dati, vengono incollati solo i valori dei dati senza gli attributi delle definizioni della variabili.
- Se si copia e si incolla una variabile intera nella Visualizzazione dei dati selezionando il nome della variabile nella parte superiore della colonna, vengono incollati tutti i dati e tutti gli attributi delle definizioni delle variabili per la variabile in questione.
- Se si copiano e si incollano gli attributi delle definizioni delle variabili o le variabili intere nella Visualizzazione variabili, vengono incollati gli attributi selezionati (o la definizione completa della variabile) ma non i valori dati.

Ridenominazione dei file di dati

Quando si apre una sorgente dati tramite i menu e le finestre di dialogo, a ciascuna sorgente viene automaticamente assegnato il nome di file dati *DataSetn*, dove *n* è il valore intero sequenziale. Quando si apre una sorgente dati tramite la sintassi del comando, non viene assegnato alcun nome di file di dati a meno che non se ne specifici uno con il comando `DATASET NAME`. Per assegnare un nome di file di dati più descrittivo:

▶ Dai menu della finestra Editor di dati relativo al file di dati di cui si vuole modificare il nome selezionare:

File
  Rinomina file di dati

▶ Immettere il nome conforme alle regole di denominazione delle variabili SPSS. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.

Capitolo
7

Preparazione dei dati

Dopo avere aperto un file di dati o immesso dei dati nell'Editor, è possibile creare report, grafici e analisi senza ulteriori operazioni preliminari. Sono tuttavia disponibili alcune opzioni aggiuntive per la preparazione dei dati che potrebbero rivelarsi utili, ad esempio:

- assegnazione delle proprietà delle variabili che descrivono i dati e determinazione delle modalità di utilizzo di determinati valori.
- Identificazione dei casi che potrebbero contenere informazioni duplicate ed esclusione di tali casi dalle analisi o eliminazione degli stessi dal file di dati.
- Creazione di nuove variabili con categorie distinte che rappresentano gli intervalli di valori delle variabili con un numero elevato di possibili valori.

Proprietà variabili

Tuttavia, i dati semplicemente immessi nella Visualizzazione dati dell'Editor dei dati oppure letti da un formato di file esterno, ad esempio un foglio elettronico di Excel o un file di dati di testo, non dispongono di alcune proprietà di variabili che potrebbero essere molto utili, quali:

- Definizione delle etichette di valore descrittive per codici numerici (per esempio, 0 = *Maschio* e 1 = *Femmina*).
- Identificazione dei codici di valori mancanti (per esempio, 99 = *Non applicabile*).
- Assegnazione del livello di misurazione (nominale, ordinale o di scala).

È possibile assegnare tutte queste ed altre proprietà di variabili nella Visualizzazione variabili dell'Editor dei dati. Per agevolare l'esecuzione di tale procedura è possibile ricorrere anche ai seguenti strumenti:

- **Definisci proprietà variabili** consente di definire le etichette dei valori descrittivi e i valori mancanti. Si tratta di uno strumento particolarmente utile per i dati categoriali con codici numerici utilizzati per i valori delle categorie.
- **Copia proprietà dei dati** consente di utilizzare un file di dati SPSS esistente come modello per le proprietà di file e variabili nel file di dati corrente. Si tratta di uno strumento particolarmente utile se si utilizzano spesso file di dati esterni con contenuto simile, come i report mensili in formato Excel.

Definizione di proprietà delle variabili

Definisci proprietà variabili semplifica il processo di creazione di etichette di valori descrittive per variabili categoriali (nominali, ordinali). Definisci proprietà variabili:

- Esamina i valori effettivi dei dati ed elenca tutti i valori univoci dei dati per ogni variabile selezionata.
- Identifica i valori senza etichetta e fornisce una funzione per l'aggiunta automatica dell'etichetta.
- Consente di copiare etichette di valori definite da un'altra variabile nella variabile selezionata o dalla variabile selezionata in più variabili aggiuntive.

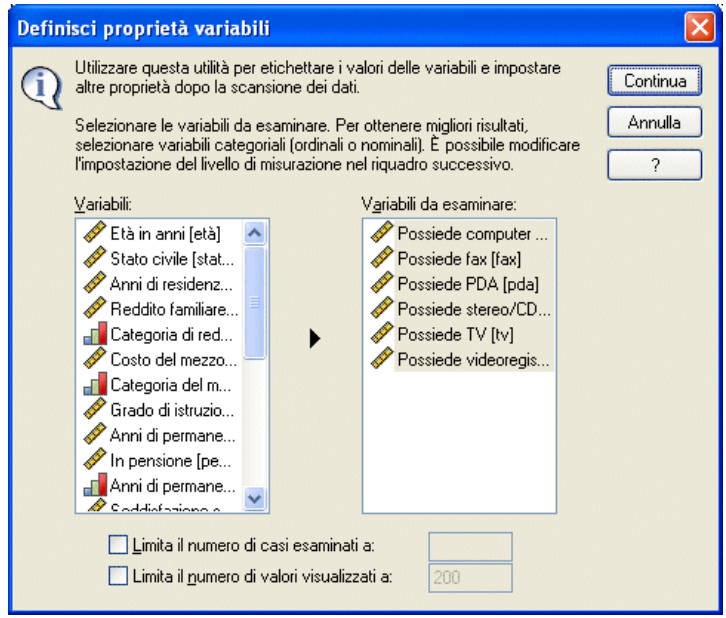
Nota: per utilizzare Definisci proprietà variabili senza eseguire la scansione dei casi, specificare 0 per il numero di casi da esaminare.

Per definire le proprietà delle variabili

▶ Dai menu, scegliere:

Dati
Definisci proprietà variabili...

Figura 7-1
Finestra di dialogo iniziale per la selezione delle variabili da definire


Definisci proprietà variabili
Utilizzare questa utilità per etichettare i valori delle variabili e impostare altre proprietà dopo la scansione dei dati.
Selezionare le variabili da esaminare. Per ottenere migliori risultati, selezionare variabili categoriali (ordinali o nominali). È possibile modificare l'impostazione del livello di misurazione nel riquadro successivo.
Continua
Annulla
?
Variabili:
Età in anni [età]
Stato civile [stat...
Anni di residenz...
Reddito familiare...
Categoria di red...
Costo del mezzo...
Categoria del m...
Grado di istruzio...
Anni di permane...
In pensione [pe...
Anni di permane...
Variabili da esaminare:
Possiede computer ...
Possiede fax [fax]
Possiede PDA [pda]
Possiede stereo/CD...
Possiede TV [tv]
Possiede videoregis...
Limita il numero di casi esaminati a:
Limita il numero di valori visualizzati a:
200

- Selezionare le variabili stringa corta o numeriche per le quali si desidera creare etichette di valori oppure definire o modificare altre proprietà delle variabili, ad esempio valori mancanti o etichette di variabile descrittive.

  Nota: Le variabili stringa lunga (variabili stringa composte da un numero di caratteri definito e maggiore di 8) non vengono visualizzate nell'elenco delle variabili. Per le variabili stringa lunga non sono disponibili etichette di valori o categorie per valori mancanti.

- Specificare il numero di casi da esaminare per generare l'elenco di valori univoci. Questo risulta particolarmente utile per i file di dati con un numero elevato di casi per cui una scansione del file di dati completo potrebbe richiedere una notevole quantità di tempo.

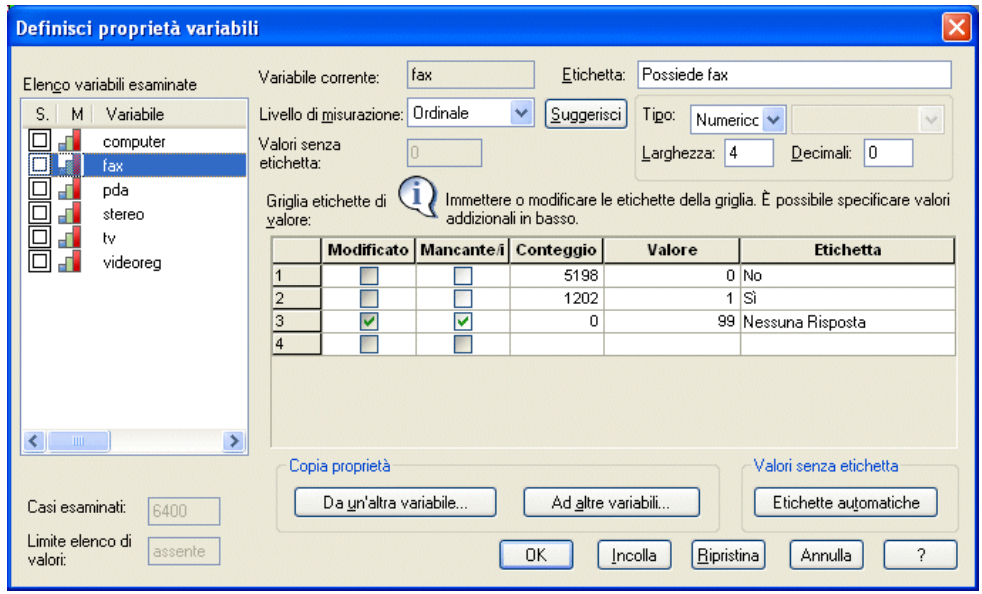
- Specificare un limite superiore per il numero di valori univoci da visualizzare. Ciò è utile soprattutto per impedire che vengano elencati centinaia, migliaia o anche milioni di valori per variabili di scala (rapporto, intervallo continuo).

- Fare clic su Continua per aprire la finestra di dialogo Definisci proprietà variabili.

- Selezionare una variabile per cui si desidera creare etichette di valori oppure definire o modificare altre proprietà delle variabili.
- Immettere il testo dell'etichetta per tutti i valori senza etichetta visualizzati nella Griglia etichette di valore.
- Se si desidera creare etichette di valore per valori non visualizzati, è possibile immettere tali valori nella colonna *Valore* che si trova al di sotto dell'ultimo valore esaminato.
- Ripetere l'operazione per ogni variabile elencata per cui si desidera creare etichette di valori.
- Fare clic su OK per applicare le etichette di valore e le altre proprietà delle variabili definite.

Definizione di etichette di valori e di altre proprietà delle variabili

Figura 7-2
Definisci proprietà variabili, finestra di dialogo principale



La finestra di dialogo principale Definisci proprietà variabili include le seguenti informazioni sulle variabili esaminate:

Elenco variabili esaminate. Un segno di spunta nella colonna *Senza etichetta* indica che la variabile contrassegnata contiene valori senza etichetta assegnata.

Per ordinare l'elenco delle variabili in modo che le variabili con valori senza etichetta vengano visualizzate all'inizio dell'elenco:

▶ Fare clic sull'intestazione della colonna *Senza etichetta* nell'Elenco variabili esaminate.

È inoltre possibile ordinare l'elenco per nome di variabile o livello di misurazione facendo clic sull'intestazione di colonna corrispondente nell'Elenco variabili esaminate.

Griglia etichette di valore

- **Etichetta.** Visualizza tutte le etichette di valore già definite. In questa colonna è possibile aggiungere o modificare etichette.
- **Valore.** Valori univoci per ogni variabile selezionata. L'elenco dei valori univoci è basato sul numero dei casi esaminati. Ad esempio, se nel file di dati sono stati esaminati solo i primi 100 casi, l'elenco conterrà solo i valori univoci presenti in questi casi. Se il file di dati è già stato ordinato in base alla variabile per cui si desidera assegnare etichette di valori, nell'elenco verrà visualizzato un numero di valori univoci notevolmente inferiore a quello dei valori effettivamente presenti nei dati.
- **Conteggio.** Numero di occorrenze di ogni valore nei casi esaminati.
- **Mancante/i.** Valori che rappresentano i dati mancanti. È possibile modificare la designazione di valori mancanti della categoria selezionando la casella di controllo. Un segno di spunta indica che la categoria è definita come categoria di valori mancanti definiti dall'utente. Se una variabile include già un intervallo di valori definiti come valori mancanti definiti dall'utente (ad esempio da 90 a 99), per questa variabile non è possibile aggiungere né eliminare categorie di valori utilizzando Definisci proprietà variabili. Per modificare le categorie di valori mancanti per le variabili contenenti intervalli di valori mancanti, è possibile utilizzare la Visualizzazione variabili dell'Editor dei dati. Per ulteriori informazioni, vedere "Valori mancanti" in Capitolo 5 a pag. 105.
- **Modificato.** Indica che è stata aggiunta o modificata un'etichetta di valore.

Nota: Se nella finestra di dialogo iniziale per il numero dei casi è stato specificato 0, la Griglia etichette di valore sarà inizialmente vuota, eccetto le eventuali etichette di valore già esistenti e/o le categorie di valori mancanti definite per la variabile selezionata. Inoltre, il pulsante Suggerisci per il livello di misurazione risulterà disattivato.

Livello di misurazione. Le etichette di valori sono estremamente utili per le variabili categoriali (nominali e ordinali). Dal momento che alcune procedure utilizzano le variabili di scala e categoriali in modo diverso, è talvolta importante assegnare il corretto livello di misurazione. Tuttavia, a tutte le nuove variabili numeriche viene assegnato un livello di misurazione predefinito. Di conseguenza, è possibile che alcune variabili che sono in realtà categoriali vengano visualizzate come variabili di scala.

Se non si è certi del corretto livello di misurazione da assegnare a una variabile, fare clic su Suggerisci.

Copia proprietà. È possibile copiare etichette di valori e altre proprietà delle variabili da un'altra variabile nella variabile selezionata o dalla variabile selezionata in una o più variabili diverse.

Valori senza etichetta. Per creare automaticamente etichette per valori senza etichetta, fare clic su Etichette automatiche.

Etichette delle variabili e formato di visualizzazione

È possibile modificare l'etichetta descrittiva di una variabile e il formato di visualizzazione.

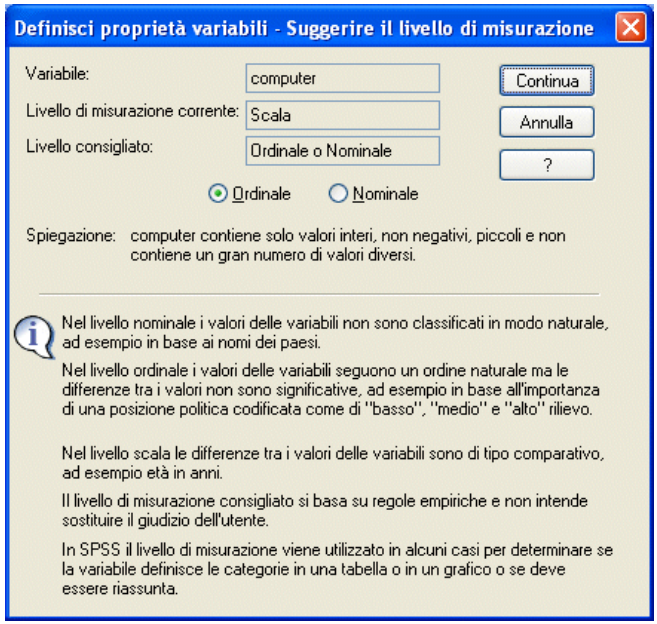
- Non è possibile modificare il tipo fondamentale della variabile (stringa o numerica).
- Per le variabili stringa, è possibile modificare solo l'etichetta e non il formato di visualizzazione.
- Per le variabili numeriche, è possibile modificare il tipo numerico (ad esempio numerico, data, dollaro o valuta personalizzata), larghezza (numero massimo di cifre, inclusi tutti i tipi di separatori di decimali e di indicatori di raggruppamento) e il numero di cifre decimali.
- Per il formato numerico di data numerico, è possibile selezionare un formato di data specifico (ad esempio gg-mm-aaaa, mm/gg/aa, aaaaggg e così via).

- Per il formato numerico personalizzato, è possibile selezionare uno dei cinque formati di valuta personalizzata, da CCA a CCE. Per ulteriori informazioni, vedere "Opzioni: Valuta" in Capitolo 44 a pag. 646.
- Nella colonna *Valore* viene visualizzato un asterisco se la larghezza specificata è minore della larghezza dei valori esaminati o dei valori visualizzati per le etichette di valori già esistenti o delle categorie di valori mancanti.
- Viene visualizzato un punto (.) se i valori esaminati o i valori visualizzati per le etichette di valori già esistenti o le categorie di valori mancanti non sono valide per il tipo di formato di visualizzazione selezionato. Ad esempio, un valore numerico interno minore di 86.400 non è valido per una variabile nel formato della data.

Assegnazione del livello di misurazione

Quando si fa clic sul pulsante Suggerisci relativo al livello di misurazione nella finestra di dialogo principale Definisci proprietà variabili, la variabile corrente viene valutata in base ai casi esaminati e alle etichette dei valori definite, quindi viene suggerito un livello di misurazione nella finestra di dialogo Suggerisci il livello di misurazione. In Spiegazione viene fornita una breve descrizione dei criteri utilizzati per stabilire il livello di misurazione consigliato.

Figura 7-3
Finestra di dialogo Suggerisci il livello di misurazione


Definisci proprietà variabili - Suggerire il livello di misurazione
Variabile:
computer
Livello di misurazione corrente:
Scala
Livello consigliato:
Ordinale o Nominale
Continua
Annulla
?
Ordinale
Nominale
Spiegazione: computer contiene solo valori interi, non negativi, piccoli e non contiene un gran numero di valori diversi.
Nel livello nominale i valori delle variabili non sono classificati in modo naturale, ad esempio in base ai nomi dei paesi.
Nel livello ordinale i valori delle variabili seguono un ordine naturale ma le differenze tra i valori non sono significative, ad esempio in base all'importanza di una posizione politica codificata come di "basso", "medio" e "alto" rilievo.
Nel livello scala le differenze tra i valori delle variabili sono di tipo comparativo, ad esempio età in anni.
Il livello di misurazione consigliato si basa su regole empiriche e non intende sostituire il giudizio dell'utente.
In SPSS il livello di misurazione viene utilizzato in alcuni casi per determinare se la variabile definisce le categorie in una tabella o in un grafico o se deve essere riassunta.

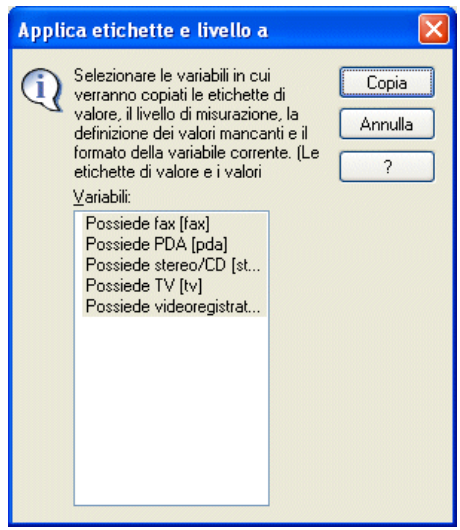
Nota: nella valutazione del livello di misurazione non sono inclusi i valori che rappresentano valori mancanti. Ad esempio, la spiegazione del livello di misurazione può indicare che il suggerimento è in parte basato sull'assenza di valori negativi nella variabile, mentre in effetti è possibile che la variabile contenga valori negativi definiti come valori mancanti.

- Fare clic su Continua per accettare il livello di misurazione suggerito o su Annulla per non accettarlo e mantenere il livello di misurazione stabilito.

Copia di proprietà delle variabili

La finestra Applica etichette e livello viene visualizzata quando si fa clic su Da un'altra variabile o Ad altre variabili nella finestra di dialogo principale Definisci proprietà variabili. Visualizza tutte le variabili esaminate corrispondenti al tipo di variabile corrente (numerico o stringa). Per le variabili stringa, è necessario che corrisponda anche la larghezza.

Figura 7-4
Finestra di dialogo Applica etichette e livello


Applica etichette e livello a
Selezionare le variabili in cui verranno copiati le etichette di valore, il livello di misurazione, la definizione dei valori mancanti e il formato della variabile corrente. (Le etichette di valore e i valori
Variabili:
Possiede fax [fax]
Possiede PDA [pda]
Possiede stereo/CD [st...
Possiede TV [tv]
Possiede videoregistrat...
Copia
Annulla
?

- Selezionare una singola variabile da cui copiare le etichette dei valori e altre proprietà della variabile (eccetto l'etichetta della variabile).

  oppure

- Selezionare una o più variabili in cui incollare le etichette dei valori e altre proprietà di una variabile.

- Fare clic su Copia per copiare le etichette dei valori e il livello di misurazione.
  - Le etichette dei valori esistenti e le categorie dei valori mancanti delle variabili di destinazione non verranno sostituite.
  - Le etichette dei valori e le categorie dei valori mancanti relative a valori che non sono ancora stati definiti nelle variabili di destinazione, vengono aggiunte all'insieme di etichette dei valori e categorie di valori mancanti di queste variabili.
  - Il livello di misurazione delle variabili di destinazione viene sempre sostituito.
  - Se la variabile sorgente o la variabile di destinazione include un intervallo di valori mancanti definito, le definizioni dei valori mancanti non vengono copiate.

Insiemi a risposta multipla

La funzione Generatore di grafici e tabelle supporta un tipo speciale di "variabile" chiamata **insieme a risposta multipla**. Gli insiemi a risposta multipla non possono essere considerati variabili nell'accezione propria del termine, poiché non è possibile visualizzarli nell'Editor dei dati e non vengono riconosciuti dalle altre procedure. Gli insiemi a risposta multipla utilizzano più variabili per registrare le risposte alle domande in cui il rispondente può fornire più di una risposta. Tali insiemi vengono considerati come variabili categoriali e nella maggior parte dei casi consentono di eseguire le stesse operazioni previste per questo tipo di variabili.

Gli insiemi a risposta multipla vengono creati utilizzando più variabili del file di dati e rappresentano un tipo di insieme speciale all'interno di un file di dati in formato SPSS. È possibile definire e salvare insiemi a risposta multipla in un file di dati in formato SPSS, ma non è possibile importare o esportare tali insiemi da o in altri formati di file. È possibile copiare gli insiemi a risposta multipla da altri file di dati SPSS tramite l'opzione Copia proprietà dati accessibile tramite il menu Dati nella finestra dell'Editor dei dati Per ulteriori informazioni, vedere "Copia delle proprietà dei dati" a pag. 136.)

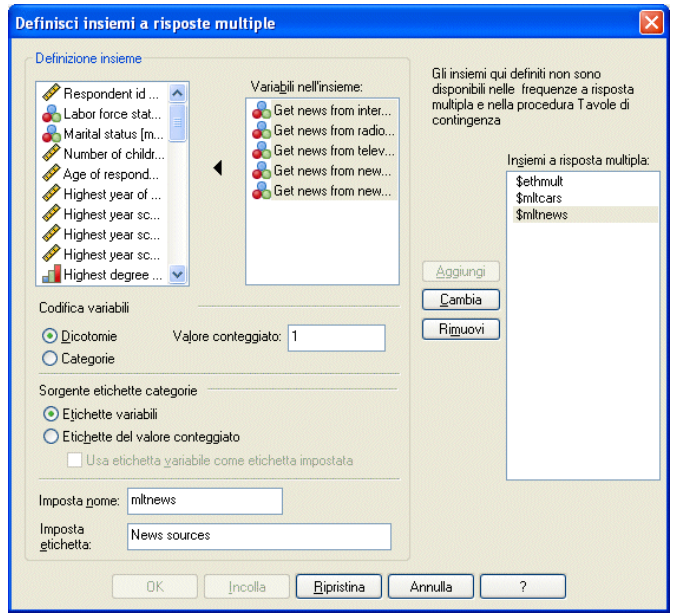
Definizione di insiemi a risposta multipla

Per definire gli insiemi a risposta multipla:

▶ Dai menu, scegliere:

Dati
  Insiemi a risposta multipla...

Figura 7-5
Finestra di dialogo Insiemi a risposta multipla



- Selezionare due o più variabili. Se le variabili sono codificate come dicotomie, indicare il valore che si desidera calcolare.
- Immettere un nome univoco per ciascun insieme a risposta multipla. Il nome può contenere fino a 63 byte. All'inizio del nome viene aggiunto automaticamente un segno di dollaro.
- Digitare un'etichetta descrittiva per l'insieme (facoltativo).
- Scegliere Aggiungi per aggiungere l'insieme a risposta multipla all'elenco di insiemi definiti.

Dicotomie (Definizione gruppi a risposta multipla)

Gli insiemi a dicotomie multiple sono in genere composti da più variabili dicotomiche, ovvero variabili che prevedono solo due possibili valori: sì/no, presente/assente, selezionato/non selezionato. Benché sia possibile che le variabili non siano effettivamente dicotomiche, tutte le variabili dell'insieme sono codificate

nello stesso modo e il valore Valore conteggiato rappresenta la condizione positivo/presente/selezionato.

Si consideri, ad esempio, un sondaggio che include la domanda "Quali delle seguenti fonti vengono utilizzate per le informazioni?" per il quale sono disponibili cinque possibili risposte. Il rispondente può specificare più risposte selezionando la casella accanto a ciascuna risposta desiderata. Le cinque risposte sono rappresentate da cinque variabili nel file di dati, codificate con 0 per la risposta *No* (non selezionato) e 1 per la risposta *Sì* (selezionato). Nell'insieme a dicotomie multiple, il valore di Valore conteggiato è 1.

Il file di dati di esempio *survey_sample.sav* (nella directory *tutorial/sample_files* della directory di installazione) contiene già tre insiemi a risposta multipla. *$mltnews* è un insieme a dicotomie multiple.

▶ Selezionare *$mltnews* nell'elenco Ins. a risposta multipla.

In questo modo vengono visualizzate le variabili e le impostazioni utilizzate per definire l'insieme a risposta multipla specificato.

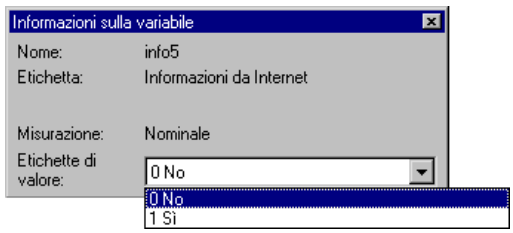
- Nell'elenco Variabili nell'insieme sono visualizzate le cinque variabili utilizzate per creare l'insieme a risposta multipla.
- Nel gruppo Variabili codificate come è indicato che le variabili sono dicotomiche.
- Il valore di Valore conteggiato è 1.

▶ Selezionare una delle variabili nell'elenco Variabili nell'insieme.

▶ Fare clic con il pulsante destro del mouse sulla variabile e scegliere Informazioni sulla variabile dal menu di scelta rapida.

▶ Nella finestra Informazioni sulla variabile fare clic sulla freccia dell'elenco a discesa Etichette di valore per visualizzare l'elenco completo delle etichette di valore definite.

Figura 7-6
Informazioni sulla variabile per la variabile sorgente delle dicotomie multiple



Le etichette di valore indicano che la variabile è dicotomica, con i valori 0 e 1 che rappresentano rispettivamente *No* e *Sì*. Tutte e cinque le variabili dell'elenco sono codificate nello stesso modo e il valore 1 (codice per *Sì*) è il valore conteggiato per l'insieme a dicotomie multiple.

Categorie

Un insieme a categorie multiple è composto da più variabili codificate nello stesso modo e spesso prevede numerose categorie possibili di risposta. Si consideri, ad esempio, un sondaggio che prevede la seguente specifica: "Indicare fino a tre nazionalità che meglio descrivono la propria identità etnica". Le risposte possibili sono centinaia, ma per motivi di codifica l'elenco è limitato a 40 nazionalità e include una categoria "Altro" che consente di specificare voci diverse. Nel file di dati, le tre scelte possibili rappresentano tre variabili, ciascuna delle quali include 41 categorie (40 nazionalità codificate e la categoria "Altro").

Nel file di dati di esempio, *$ethmult* e *$mltcars* sono insiemi a categorie multiple.

Sorgente etichetta categoria

Per le dicotomie multiple è possibile controllare la modalità di etichettatura degli insiemi.

- **Etichette di variabili.** Le etichette di variabili definite (o nomi delle variabili per le variabili prive di etichette di variabili definite) vengono utilizzate come etichette di categoria dell'insieme. Ad esempio, se tutte le variabili dell'insieme possiedono la stessa etichetta di valore (o nessuna etichetta di valore definita) per il valore conteggiato (ad esempio, *Sì*), è possibile utilizzare le etichette di variabili come etichette di categoria dell'insieme.
- **Etichette dei valori conteggiati.** Le etichette di valore definite dei valori conteggiati vengono utilizzate come etichette di categoria dell'insieme. Selezionare questa opzione solo se tutte le variabili possiedono un'etichetta di valore definita per il valore conteggiato e tale etichetta è diversa per ogni variabile.
- **Utilizzare l'etichetta di variabile come etichetta dell'insieme.** Se si seleziona Etichette dei valori conteggiati, è inoltre possibile utilizzare l'etichetta di variabile per la prima variabile dell'insieme che possiede un'etichetta di variabile definita come etichetta dell'insieme. Se nessuna delle variabili dell'insieme possiede un'etichetta di variabile definita, il nome della prima variabile dell'insieme viene utilizzata come etichetta dell'insieme stesso.

Copia delle proprietà dei dati

La Copia proprietà dei dati guidata consente di utilizzare un file dati SPSS esterno come modello per definire le proprietà di file e variabili nel file dati attivo. È possibile anche usare le variabile nel file di dati attivo come modelli per altre variabili dello stesso file. È possibile:

- Copiare le proprietà del file selezionate da un file di dati esterno o da un file di dati aperto nel file di dati attivo. Le proprietà dei file includono documenti, etichette dei file, insiemi a risposta multipla, insiemi di variabili e ponderazione.
- Copiare le proprietà della variabile selezionata da un file di dati esterno da un file di dati aperto che corrisponda alle variabili al file di dati attivo. Le proprietà delle variabili includono etichette dei valori, valori mancanti, livello di misurazione, etichette delle variabili, formati di stampa e di scrittura, allineamento e larghezza delle colonne (nell'Editor dei dati).
- Copiare le proprietà delle variabili selezionate di una variabile di un file esterno, un file di dati aperto o un file di dati attivo in più variabili del file di dati attivo.
- Creare nuove variabili nel file di dati attivo in base alle variabili selezionate in un file di dati esterno o aperto.

Quando si copiano le proprietà dei dati, è necessario seguire le seguenti regole:

- Se come file di dati sorgente si utilizza un file di dati esterno, questo deve essere in formato SPSS.
- Se come file di dati sorgente si utilizza il file di dati attivo, questo deve includere almeno una variabile. Non è possibile usare file di dati attivi completamente vuoti come file di dati sorgente.
- Le proprietà non definite (vuote) nel file di dati sorgente non sovrascrivono le proprietà definite nel file di dati attivo.
- Le proprietà delle variabili vengono copiate dalla variabile sorgente solo nelle variabili di destinazione di tipo corrispondente, stringa (alfanumerica) o numerica (numerica, data e valuta).

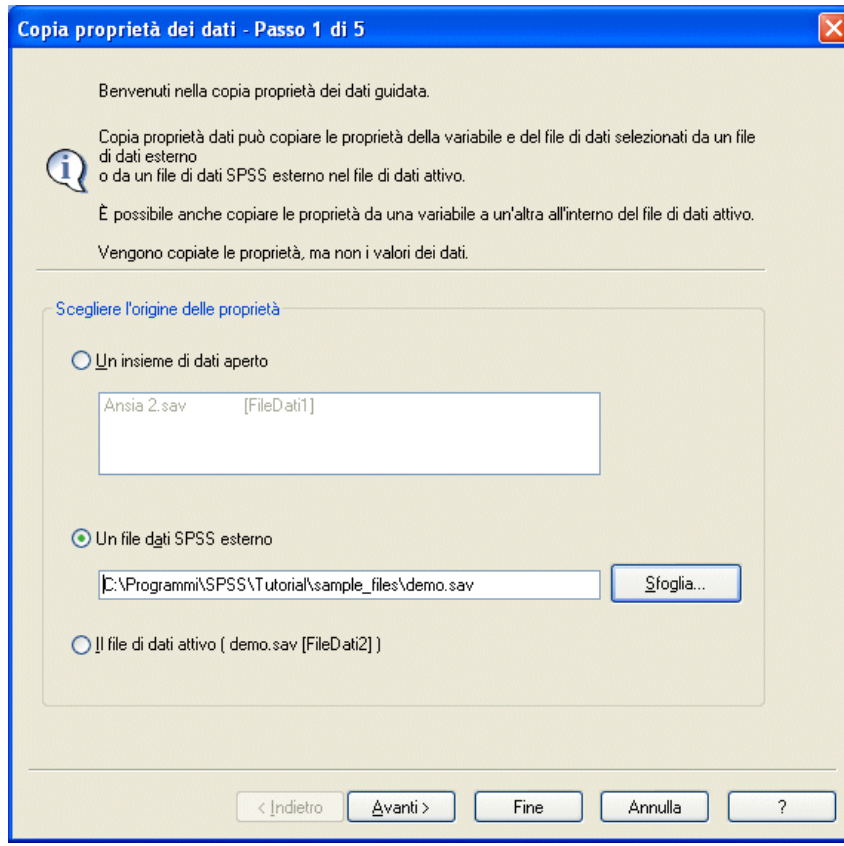
Nota: La funzione Copia proprietà dei dati sostituisce la funzione Applica dizionario dati, precedentemente disponibile nel menu File.

Per copiare le proprietà dei dati

- Dai menu nella finestra dell'Editor dei dati, scegliere:

  Dati
  Copia proprietà dei dati...

Figura 7-7
Copia proprietà dei dati guidata: Passo 1


Copia proprietà dei dati - Passo 1 di 5
Benvenuti nella copia proprietà dei dati guidata.
Copia proprietà dati può copiare le proprietà della variabile e del file di dati selezionati da un file di dati esterno
o da un file di dati SPSS esterno nel file di dati attivo.
È possibile anche copiare le proprietà da una variabile a un'altra all'interno del file di dati attivo.
Vengono copiate le proprietà, ma non i valori dei dati.
Scegliere l'origine delle proprietà
Un insieme di dati aperto
Ansia 2.sav [FileDati1]
Un file dati SPSS esterno
C:\Programmi\SPSS\Tutorial\sample_files\demo.sav
Sfoglia...
Il file di dati attivo (demo.sav [FileDati2])
< Indietro
Avanti >
Fine
Annulla
?

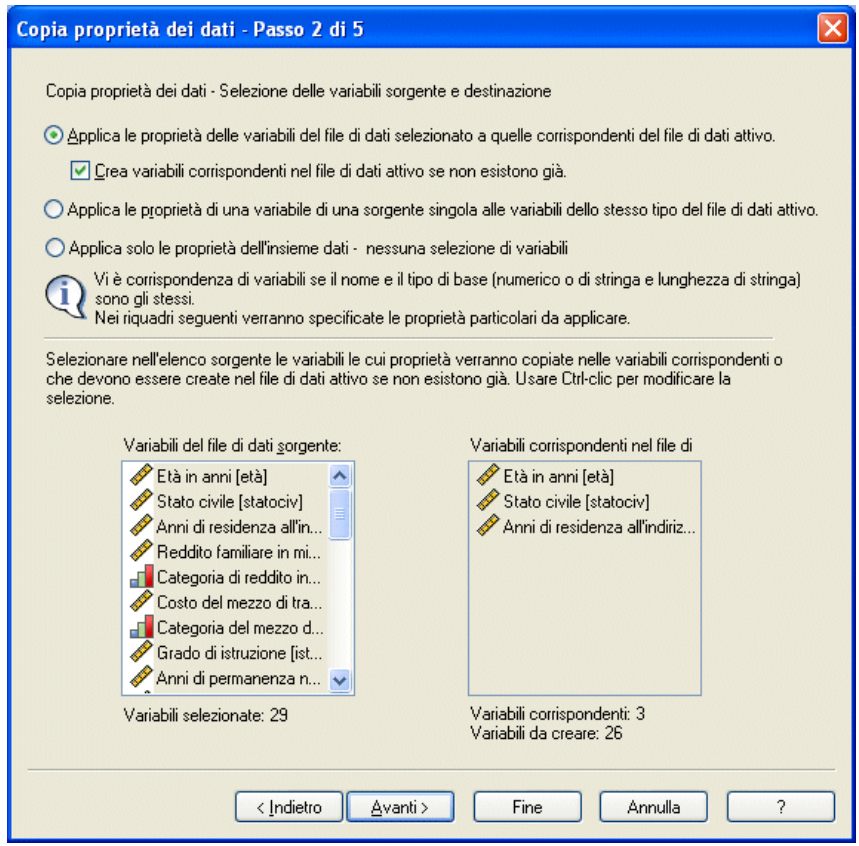
- Selezionare il file di dati e/o le proprietà della variabile da copiare. Il file può essere il file di dati correntemente aperto, un file di dati esterno in formato SPSS o un file di dati attivo.

- Eseguire le istruzioni visualizzate nella procedura guidata per la copia delle proprietà dei dati.

Selezione delle variabili sorgente e destinazione

In questo passaggio è possibile specificare le variabili sorgente che includono le proprietà delle variabili che si desidera copiare e le variabili di destinazione che riceveranno le proprietà in questione.

Figura 7-8
Copia proprietà dei dati guidata: Passo 2


Copia proprietà dei dati - Passo 2 di 5
Copia proprietà dei dati - Selezione delle variabili sorgente e destinazione
Applica le proprietà delle variabili del file di dati selezionato a quelle corrispondenti del file di dati attivo.
Crea variabili corrispondenti nel file di dati attivo se non esistono già.
Applica le proprietà di una variabile di una sorgente singola alle variabili dello stesso tipo del file di dati attivo.
Applica solo le proprietà dell'insieme dati - nessuna selezione di variabili
Vi è corrispondenza di variabili se il nome e il tipo di base (numerico o di stringa e lunghezza di stringa) sono gli stessi.
Nei riquadri seguenti verranno specificate le proprietà particolari da applicare.
Selezionare nell'elenco sorgente le variabili le cui proprietà verranno copiate nelle variabili corrispondenti o che devono essere create nel file di dati attivo se non esistono già. Usare Ctrl-clic per modificare la selezione.
Variabili del file di dati sorgente:
Età in anni [età]
Stato civile [statociv]
Anni di residenza all'in...
Reddito familiare in mi...
Categoria di reddito in...
Costo del mezzo di tra...
Categoria del mezzo d...
Grado di istruzione [ist...
Anni di permanenza n...
Variabili selezionate: 29
Variabili corrispondenti nel file di
Età in anni [età]
Stato civile [statociv]
Anni di residenza all'indiriz...
Variabili corrispondenti: 3
Variabili da creare: 26
< Indietro
Avanti >
Fine
Annulla
?

Applicare le proprietà delle variabili del file sorgente selezionato alle variabili corrispondenti del file di dati attivo. Le proprietà delle variabili vengono copiate da una o più variabili sorgente selezionate nelle variabili corrispondenti del file di dati attivo. Le variabili "corrispondono" quando condividono lo stesso nome e tipo di variabile (stringa o numerica). Per le variabili stringa, deve corrispondere anche la

lunghezza definita. Per impostazione predefinita, nei due elenchi di variabili vengono visualizzate solo le variabili corrispondenti.

- **Creare variabili corrispondenti nel file di dati attivo se non esistono già.** L'elenco sorgente verrà aggiornato in modo da includere tutte le variabili del file di dati sorgente. Se si selezionano variabili sorgente che non esistono nel file di dati attivo (in base al nome della variabile), in tale file verranno create nuove variabili con nomi e proprietà corrispondenti a quelli delle variabili del file di dati sorgente.

Se il file di dati attivo non include variabili (nuovo file di dati vuoto), vengono visualizzate tutte le variabili del file di dati sorgente e nel file di dati attivo vengono create nuove variabili in base alle variabili sorgente selezionate.

Applicare le proprietà di una singola variabile sorgente alle variabili dello stesso tipo selezionate nel file di dati attivo. È possibile applicare le proprietà delle variabili di una singola variabile selezionata nell'elenco sorgente a una o più variabili selezionate nell'elenco del file di dati attivo. Nell'elenco del file di dati attivo vengono visualizzate solo variabili dello stesso tipo (numeriche o stringa) della variabile selezionata nell'elenco sorgente. Per le variabili stringa, vengono visualizzate solo le stringhe che presentano la stessa lunghezza definita della variabile sorgente. Questa opzione non è disponibile se il file di dati attivo non contiene variabili.

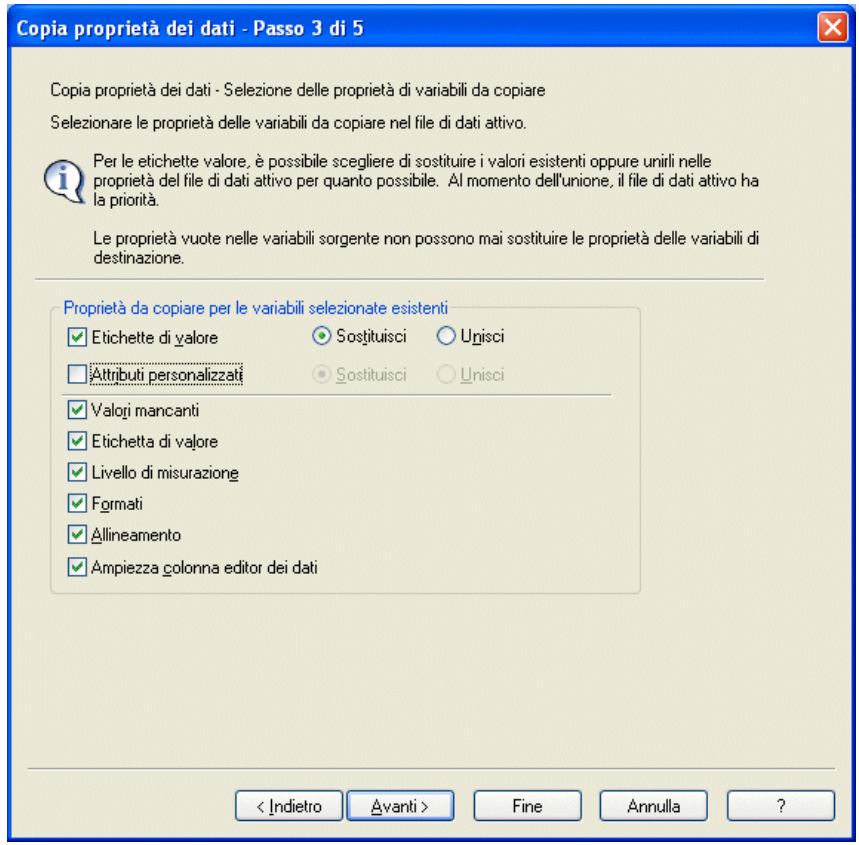
Nota: non è possibile creare nuove variabili nel file di dati attivo con questa opzione.

Applicare solo variabili del set di dati, senza selezione di variabili. Al file di dati attivo verranno applicate solo le proprietà dei file (ad esempio, documenti, etichette di file, peso). Non verranno applicate proprietà delle variabili. Questa opzione non è disponibile se il file di dati attivo è anche la sorgente del file di dati stesso.

Selezione delle proprietà delle variabili da copiare

È possibile copiare le proprietà delle variabili selezionate dalle variabili sorgente alle variabili di destinazione. Le proprietà non definite (vuote) delle variabili sorgente non sovrascrivono le proprietà definite delle variabili di destinazione.

Figura 7-9
Copia proprietà dei dati guidata: Passo 3


Copia proprietà dei dati - Passo 3 di 5
Copia proprietà dei dati - Selezione delle proprietà di variabili da copiare
Selezionare le proprietà delle variabili da copiare nel file di dati attivo.
Per le etichette valore, è possibile scegliere di sostituire i valori esistenti oppure unirli nelle proprietà del file di dati attivo per quanto possibile. Al momento dell'unione, il file di dati attivo ha la priorità.
Le proprietà vuote nelle variabili sorgente non possono mai sostituire le proprietà delle variabili di destinazione.
Proprietà da copiare per le variabili selezionate esistenti
Etichette di valore
Sostituisci
Unisci
Attributi personalizzati
Sostituisci
Unisci
Valori mancanti
Etichetta di valore
Livello di misurazione
Formati
Allineamento
Ampiezza colonna editor dei dati
< Indietro
Avanti >
Fine
Annulla
?

Etichette dei valori. Le etichette dei valori sono etichette descrittive associate ai valori dei dati. Le etichette dei valori vengono utilizzate frequentemente per rappresentare con valori di dati numerici categorie non numeriche (ad esempio, i codici 1 e 2 per *Maschile* e *Femminile*). È possibile sostituire o unire le etichette dei valori nelle variabili di destinazione.

- Il comando Sostituisci consente di eliminare tutte le etichette dei valori definite per la variabile di destinazione e di sostituirle con le etichette dei valori definite per la variabile sorgente.
- Il comando Unisci consente di unire tutte le etichette dei valori definite per la variabile sorgente alle etichette dei valori definite per la variabile di destinazione. Se allo stesso valore corrisponde un'etichetta di valore definita nelle variabili sia sorgente sia di destinazione, l'etichetta del valore nella variabile di destinazione rimarrà invariata.

Attributi personalizzati. Attributi variabili personalizzati, generalmente creati dal comando `VARIABLE ATTRIBUTE` nella sintassi dei comandi.

- Il comando Sostituisci consente di eliminare gli attributi personalizzati per la variabile di destinazione e di sostituirle con gli attributi definiti nella variabille sorgente.
- Il comando Unisci consente di unire gli attributi definiti per la variabile sorgente agli attributi definiti esistenti definiti per la variabile di destinazione.

Valori mancanti. I valori mancanti sono valori utilizzati per rappresentare dati mancanti (ad esempio, 98 per *Non so* e 99 per *Non applicabile*). In genere a questi valori sono associate etichette definite che indicano a cosa si riferiscono i codici dei valori mancanti. Tutti i valori mancanti definiti per la variabile di destinazione vengono eliminati e sostituiti con i valori mancanti definiti per la variabile sorgente.

Etichette delle variabili. Le etichette delle variabili descrittive possono contenere spazi e caratteri riservati non consentiti nei nomi delle variabili. Se si copiano le proprietà delle variabili da una singola variabile sorgente in più variabili di destinazione, è necessario valutare accuratamente l'opportunità di utilizzare questa opzione.

Livello di misurazione. Il livello di misurazione può essere nominale, ordinale o di scala. Nelle procedure in cui è rilevante la distinzione tra i diversi livelli di misurazione, i livelli nominale e ordinale vengono entrambi considerati **categoriali**.

Formati. Per le variabili numeriche, questa opzione controlla il tipo di numeri (ad esempio numerico, data o valuta), la larghezza (numero totale di caratteri visualizzati, inclusi i caratteri di testa e di coda e l'indicatore decimale) e il numero di cifre decimali visualizzate. Questa opzione viene ignorata per le variabili stringa.

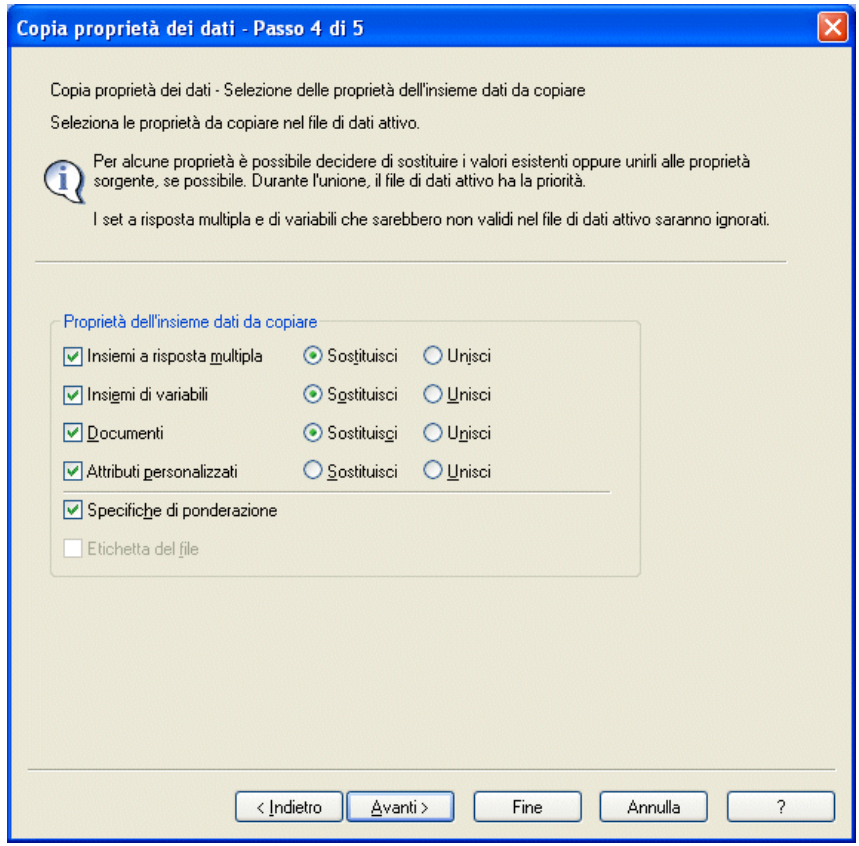
Allineamento. Questa opzione influisce solo sull'allineamento (a sinistra, a destra, al centro) nella Visualizzazione dati dell'Editor dei dati.

Larghezza delle colonne nell'Editor dei dati. Questa opzione influisce solo sulla larghezza delle colonne nella Visualizzazione dati dell'Editor dei dati.

Copia delle proprietà dell'insieme di dati (file)

È possibile applicare al file di dati attivo le proprietà globali dell'insieme di dati selezionate nel file di dati sorgente. Questa opzione non è disponibile se il file di dati attivo è anche la sorgente del file di dati stesso.

Figura 7-10
Copia proprietà dei dati guidata: Passo 4


Copia proprietà dei dati - Passo 4 di 5
Copia proprietà dei dati - Selezione delle proprietà dell'insieme dati da copiare
Seleziona le proprietà da copiare nel file di dati attivo.
Per alcune proprietà è possibile decidere di sostituire i valori esistenti oppure unirli alle proprietà sorgente, se possibile. Durante l'unione, il file di dati attivo ha la priorità.
I set a risposta multipla e di variabili che sarebbero non validi nel file di dati attivo saranno ignorati.
Proprietà dell'insieme dati da copiare
Insiemi a risposta multipla
Sostituisci
Unisci
Insiemi di variabili
Sostituisci
Unisci
Documenti
Sostituisci
Unisci
Attributi personalizzati
Sostituisci
Unisci
Specifiche di ponderazione
Etichetta del file
< Indietro
Avanti >
Fine
Annulla
?

Insiemi a risposta multipla. Applica al file di dati attivo le definizioni dell'insieme a risposta multipla del file di dati sorgente. (*Nota*: gli insiemi a risposta multipla vengono utilizzati solo dal Generatore di Grafici e dal componente aggiuntivo Tabelle).

- Gli insiemi a risposta multipla del file di dati sorgente che includono variabili che non esistono nel file di dati attivo vengono ignorati, a meno che non vengano create le variabili necessarie in base alle indicazioni fornite nel passo 2 (Selezione delle variabili sorgente e di destinazione) della Copia proprietà dei dati guidata.
- Il comando Sostituisci consente di eliminare tutti gli insiemi a risposta multipla inclusi nel file di dati attivo e di sostituirli con gli insiemi a risposta multipla del file di dati sorgente.
- Il comando Unisci consente di unire gli insiemi a risposta multipla del file di dati di origine al gruppo di insiemi a risposta multipla del file di dati attivo. Se entrambi i file includono un insieme con lo stesso nome, l'insieme del file di dati attivo rimarrà invariato.

Insiemi di variabili. Gli insiemi di variabili vengono utilizzati per controllare l'elenco delle variabili visualizzate nelle finestre di dialogo. Gli insiemi di variabili vengono definiti scegliendo Definisci insiemi dal menu Strumenti.

- Gli insiemi del file di dati sorgente che includono variabili che non esistono nel file di dati attivo verranno ignorati, a meno che tali variabili non vengano create in base alle indicazioni del passo 2 (Selezione di variabili sorgente e di destinazione) della Copia proprietà dei dati guidata.
- Il comando Sostituisci consente di eliminare gli insiemi di variabili esistenti nel file di dati attivo e di sostituirli con gli insiemi di variabili del file di dati sorgente.
- Il comando Unisci consente di aggiungere insiemi di variabili del file di dati sorgente al gruppo di insiemi di variabili del file di dati attivo. Se entrambi i file includono un insieme con lo stesso nome, l'insieme del file di dati attivo rimarrà invariato.

Documenti. Note aggiunte al file di dati mediante il comando `DOCUMENT`.

- Il comando Sostituisci consente di eliminare i documenti esistenti nel file di dati attivo e di sostituirli con i documenti del file di dati sorgente.
- Il comando Unisci consente di combinare i documenti dei file di dati sorgente e attivo. Al file di dati attivo vengono aggiunti i documenti univoci del file di dati sorgente che non esistono al suo interno. Tutti i documenti vengono quindi ordinati in base alla data.

Attributi personalizzati. Attributi dei file di dati personalizzati, generalmente creati dal comando `DATAFILE ATTRIBUTE` nella sintassi dei comandi.

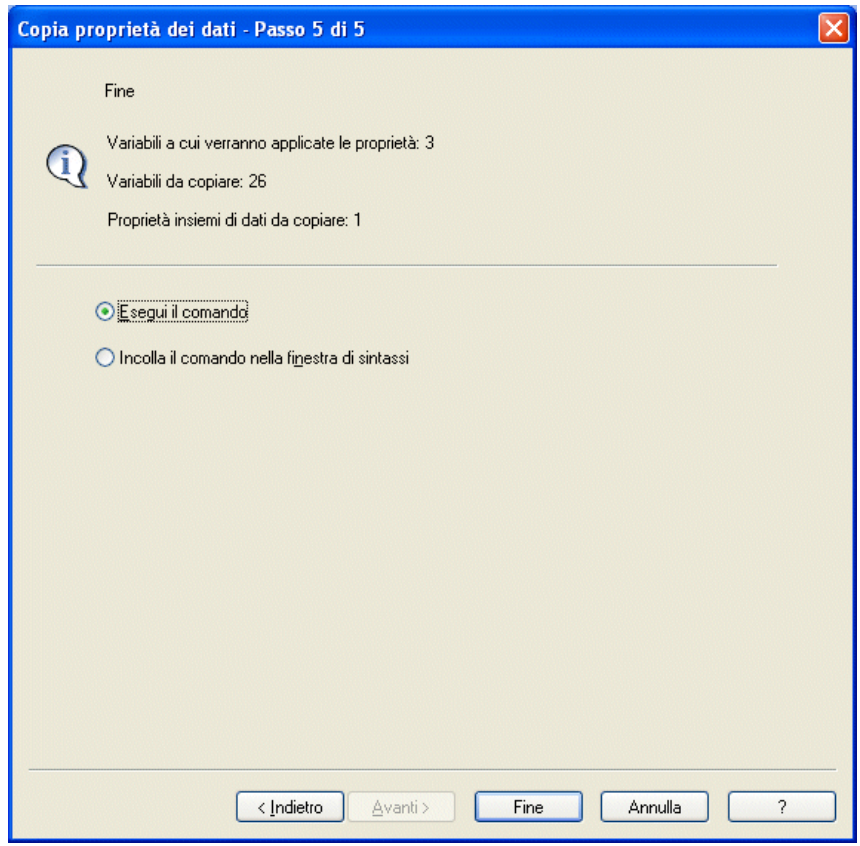
- Sostituisci elimina tutti gli attributi dei file di dati personalizzati esistenti nel file di dati attivo, sostituendoli con quelli del file di dati sorgente.
- Il comando Unisci consente di combinare gli attributi del file di dati dei file di dati sorgente e attivo. Al file di dati attivo vengono aggiunti i nomi degli attributi univoci del file di dati sorgente che non esistono al suo interno. Se il nome dell'attributo è già presente in entrambi i file di dati, l'attributo nominato nel file di dati attivo rimane invariato.

Specifiche di ponderazione. Esegue la ponderazione dei casi in base alla variabile di ponderazione corrente del file di dati sorgente, se esiste una variabile corrispondente nel file di dati attivo. Questa operazione ha la priorità rispetto a qualsiasi processo di ponderazione in atto nel file di dati attivo.

Etichetta del file. Etichetta descrittiva applicata a un file di dati mediante il comando `FILE LABEL`.

Risultati

Figura 7-11
Copia proprietà dei dati guidata: Passo 5


Copia proprietà dei dati - Passo 5 di 5
Fine
Variabili a cui verranno applicate le proprietà: 3
Variabili da copiare: 26
Proprietà insiemi di dati da copiare: 1
Esegui il comando
Incolla il comando nella finestra di sintassi
< Indietro
Avanti >
Fine
Annulla
?

Nell'ultimo passo della Copia proprietà dei dati guidata vengono fornite informazioni sul numero di variabili per cui verranno copiate proprietà dal file di dati sorgente, sul numero di variabili che verranno create e sul numero di proprietà dell'insieme di dati (file) che verranno copiate.

È inoltre possibile incollare la sintassi dei comandi generata in una finestra di sintassi e salvarla per un successivo utilizzo.

Identificazione di casi duplicati

La presenza di casi duplicati nei dati può essere dovuta a svariati motivi, ad esempio:

- Errori nell'inserimento dei dati, che comportano l'inserimento accidentale dello stesso caso più volte.
- Più casi che condividono lo stesso ID principale ma con ID secondari diversi, ad esempio i membri di una famiglia che risiedono nella stessa abitazione.
- Più casi che rappresentano lo stesso caso, ma per i quali sono stati impostati valori diversi per le variabili diverse dalle variabili di identificazione del caso, ad esempio nel caso di più acquisti effettuati dalla stessa persona o società in relazione a prodotti diversi o in momenti diversi.

L'opzione Identifica i casi duplicati consente di definire i *duplicati* nel modo desiderato e di gestire la determinazione automatica dei casi principali e dei casi duplicati.

Per identificare e contrassegnare i casi duplicati:

▶ Dai menu, scegliere:

Dati
Identifica i casi duplicati...

▶ Selezionare una o più variabili che identificano i casi corrispondenti.

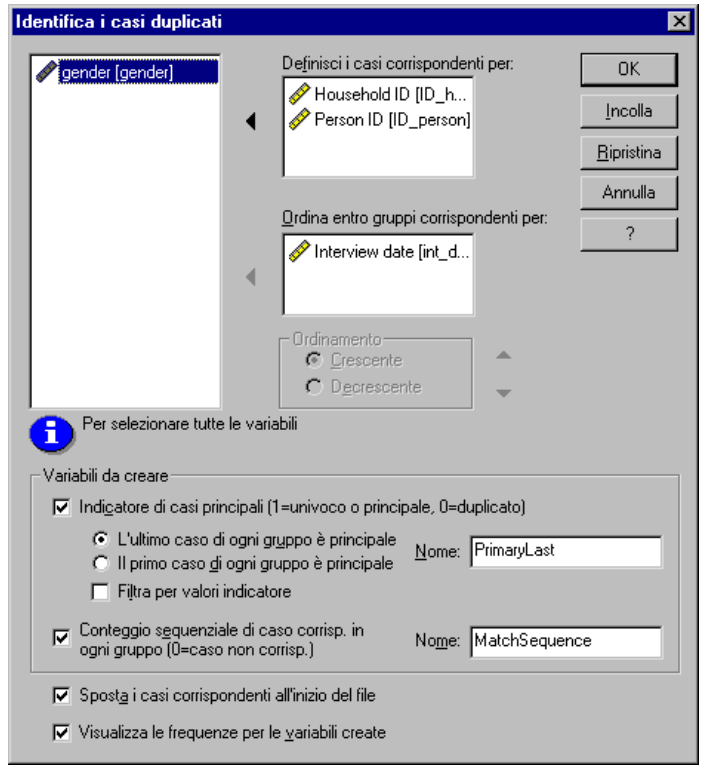
▶ Selezionare una o più opzioni del gruppo Variabili da creare.

Se lo si desidera, è possibile:

▶ Selezionare una o più variabili per ordinare i casi entro i gruppi definiti dalle variabili dei casi corrispondenti selezionate. L'ordinamento definito da tali variabili determina il primo e l'ultimo caso di ogni gruppo. In caso contrario, verrà utilizzato l'ordinamento del file originale.

▶ Filtrare automaticamente i casi duplicati in modo tale che non vengano inseriti nei rapporti, nei grafici o nel calcolo delle statistiche.

Figura 7-12
Finestra di dialogo Identifica i casi duplicati



Definisci i casi corrispondenti per. I casi vengono considerati come duplicati se i relativi valori per *tutte* le variabili selezionate risultano corrispondenti. Se si desidera identificare solo i casi corrispondenti al 100%, selezionare tutte le variabili.

Ordina entro gruppi corrispondenti per. I casi vengono ordinati automaticamente in base alle variabili che definiscono i casi corrispondenti. È possibile selezionare variabili di ordinamento aggiuntive in base alle quali verrà determinato l'ordine sequenziale dei casi in ogni gruppo corrispondente.

- Per ogni variabile di ordinamento è possibile eseguire l'ordinamento crescente o decrescente.

- Se si selezionano più variabili di ordinamento, i casi verranno ordinati in base a ciascuna variabile all'interno delle categorie della variabile precedente nell'elenco. Se, ad esempio, si seleziona *data* come prima variabile di ordinamento e *quantità* come seconda variabile di ordinamento, i casi verranno classificati per quantità all'interno di ciascuna categoria di data.
- Per modificare l'ordine delle variabili, utilizzare le frecce in su e in giù a destra dell'elenco.
- L'ordinamento determina il primo e l'ultimo caso all'interno di ogni gruppo corrispondente, che a sua volta determina il valore della variabile indicatore principale facoltativa. Se, ad esempio, si desidera filtrare tutti i casi tranne il più recente di ogni gruppo corrispondente, è possibile ordinare i casi all'interno del gruppo in ordine crescente in base a una variabile di data. In questo modo, la data più recente risulterà essere l'ultima data nel gruppo.

Indicatore di casi principali. Crea una variabile con il valore 1 per tutti i casi univoci e il caso identificato come il caso principale in ogni gruppo di casi corrispondenti e con un valore 0 per i duplicati non principali in ogni gruppo.

- Il caso principale può essere l'ultimo o il primo caso di ogni gruppo corrispondente, in base all'ordinamento all'interno del gruppo corrispondente. Se non si specificano variabili di ordinamento, l'ordine dei casi entro ogni gruppo sarà determinato dall'ordinamento del file originale.
- È possibile utilizzare la variabile indicatore come una **variabile filtro** per escludere i duplicati non principali dai rapporti e dalle analisi senza eliminare i casi dal file di dati.

Conteggio sequenziale di caso corrisp. in ogni gruppo. Crea una variabile con un valore sequenziale compreso tra 1 e *n* per i casi di ogni gruppo corrispondente. La sequenza si basa sull'ordine corrente dei casi in ogni gruppo, che è rappresentato dall'ordinamento del file originale oppure dall'ordine determinato dalle variabili di ordinamento specificate.

Sposta i casi corrispondenti all'inizio del file. Ordina il file dati in modo tale che tutti i gruppi di casi corrispondenti si trovino all'inizio del file, semplificando il controllo visuale dei casi corrispondenti nell'Editor dei dati.

Visualizza le frequenze per la variabili create. Tabelle di frequenza contenenti i conteggi relativi a ogni valore delle variabili create. Ad esempio, per la variabile indicatore principale, nella tabella verrà visualizzato il numero di casi con il valore 0

per la variabile, a indicare il numero di duplicati, e il numero di casi con il valore 1, a indicare il numero di casi univoci e principali.

Valori mancanti. Per le variabili numeriche, il valore mancante di sistema viene considerato come qualsiasi altro valore.—I casi con un valore mancante di sistema per una variabile identificatrice vengono considerati come casi che hanno valori corrispondenti per la variabile specificata. Per le variabili stringa, i casi senza valori per una variabile identificatrice vengono considerati come casi che hanno valori corrispondenti per la variabile.

Segmentazione grafica

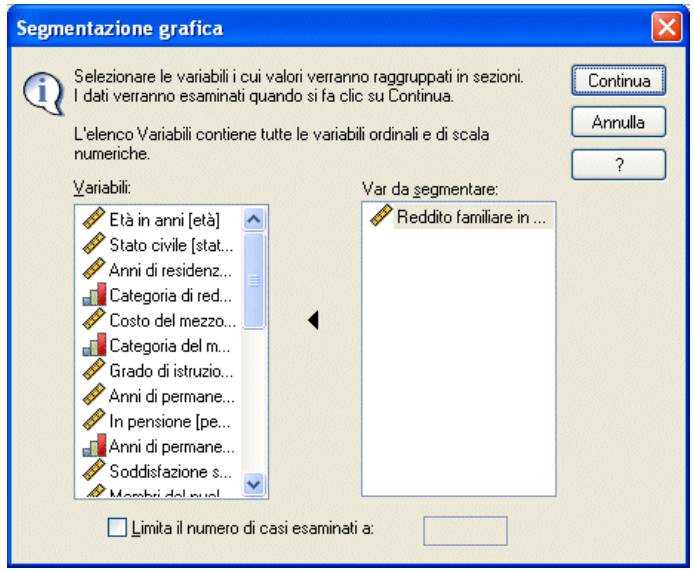
L'utilità Segmentazione grafica è progettata per semplificare il processo di creazione di nuove variabili basate sul raggruppamento di valori contigui di variabili esistenti in un numero limitato di categorie distinte. È possibile utilizzare Segmentazione grafica per:

- Creare variabili categoriali da variabili di scala continue. Ad esempio, è possibile utilizzare una variabile continua relativa al reddito per creare una nuova variabile categoriale contenente intervalli di reddito.
- Comprimere un elevato numero di categorie ordinali in un insieme più piccolo di categorie. Ad esempio, è possibile comprimere una scala di stime di nove categorie in tre categorie che rappresentino i valori bassi, medi ed elevati.

Nel primo passo di Segmentazione grafica, è possibile:

▶ Selezionare la scala numerica e/o le variabili ordinali per cui si desidera creare le nuove variabili categoriali (segmentate).

Figura 7-13
Finestra di dialogo iniziale per la selezione delle variabili da segmentare


Segmentazione grafica
Selezionare le variabili i cui valori verranno raggruppati in sezioni. I dati verranno esaminati quando si fa clic su Continua.
L'elenco Variabili contiene tutte le variabili ordinali e di scala numeriche.
Continua
Annulla
?
Variabili:
Età in anni [età]
Stato civile [stat...
Anni di residenz...
Categoria di red...
Costo del mezzo...
Categoria del m...
Grado di istruzio...
Anni di permane...
In pensione [pe...
Anni di permane...
Soddisfazione s...
Var da segmentare:
Reddito familiare in ...
Limita il numero di casi esaminati a:

È inoltre possibile limitare il numero di casi da esaminare. Per file di dati con un elevato numero di casi, la limitazione del numero di casi esaminati consente di risparmiare tempo, tuttavia è consigliabile evitarla, se possibile, perché influenza la distribuzione dei valori utilizzati per calcoli successivi in Segmentazione grafica.

Nota: le variabili stringa e le variabili nominali numeriche non vengono visualizzate nell'elenco di variabili sorgente. Segmentazione grafica richiede variabili numeriche, misurate su una scala o un livello ordinale, in quanto presume che i valori dei dati rappresentano un tipo di ordine logico che può essere utilizzato per raggruppare valori in modo significativo. È possibile modificare il livello di misurazione definito per una variabile nella Visualizzazione variabili dell'Editor dei dati. Per ulteriori informazioni, vedere "Livello di misurazione delle variabili" in Capitolo 5 a pag. 99.

Per segmentare le variabili

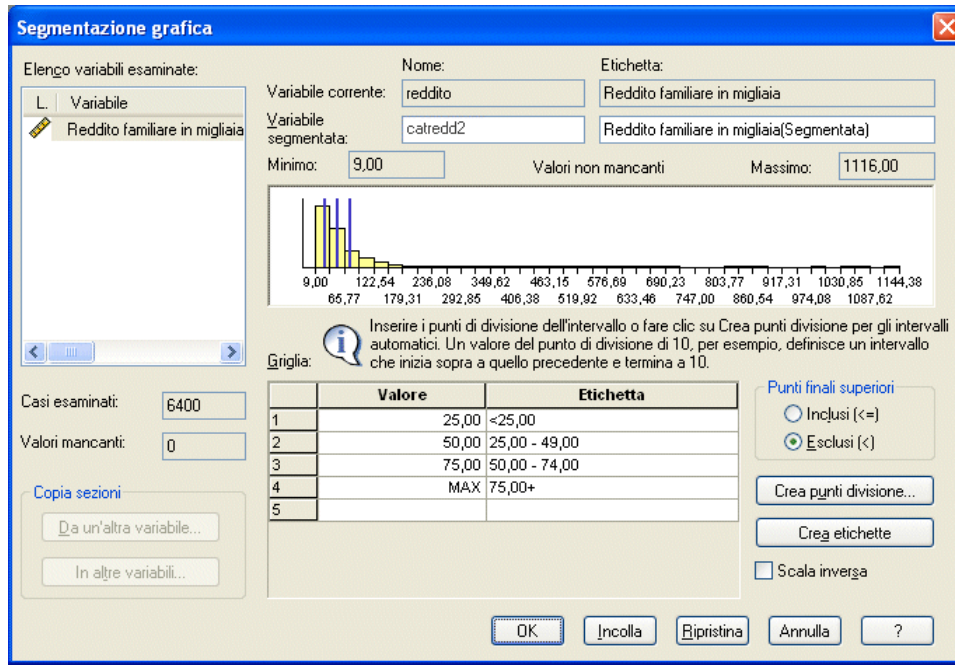
- Dai menu nella finestra dell'Editor dei dati, scegliere:

  Trasforma
  Segmentazione grafica...

- Selezionare la scala numerica e/o le variabili ordinali per cui si desidera creare le nuove variabili categoriali (segmentate).
- Selezionare una variabile dall'Elenco variabili esaminate.
- Specificare un nome per la nuova variabile segmentata. I nomi delle variabili devono essere univoci e rispettare le regole di denominazione delle variabili SPSS. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.
- Definire il criterio di segmentazione per la nuova variabile. Per ulteriori informazioni, vedere "Segmentazione delle variabili" a pag. 151.
- Fare clic su OK.

Segmentazione delle variabili

Figura 7-14
Finestra di dialogo principale di Segmentazione grafica



La finestra di dialogo principale di Segmentazione grafica include le seguenti informazioni sulle variabili esaminate:

Elenco variabili esaminate. Visualizza le variabili selezionate nella finestra di dialogo iniziale. È possibile ordinare l'elenco in base al livello di misurazione (di scala o ordinale) oppure in base al nome o all'etichetta delle variabili facendo clic sulle intestazioni delle relative colonne.

Casi esaminati. Indica il numero di casi esaminati. Tutti i casi esaminati privi di valori mancanti di sistema o di valori mancanti definiti dall'utente per la variabile selezionata vengono utilizzati per generare la distribuzione dei valori utilizzati per i calcoli in Segmentazione grafica, incluso l'istogramma visualizzato nella finestra di dialogo principale e i punti di divisione basati su percentili o unità di deviazione standard.

Valori mancanti. Indica il numero di casi esaminati con valori mancanti di sistema o valori mancanti definiti dall'utente. I valori mancanti vengono esclusi da tutte le categorie segmentate. Per ulteriori informazioni, vedere "Valori mancanti definiti dall'utente in Segmentazione grafica" a pag. 159.

Variabile corrente. Il nome e l'eventuale etichetta della variabile attualmente selezionata che verrà utilizzata come base della nuova variabile segmentata.

Var. segmentata. Nome ed etichetta facoltativa della nuova variabile segmentata.

- **Nome.** È necessario specificare un nome per la nuova variabile. I nomi delle variabili devono essere univoci e rispettare le regole di denominazione delle variabili SPSS. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.
- **Etichetta.** È possibile specificare un'etichetta di variabile descrittiva contenente un massimo di 255 caratteri. L'etichetta di variabile predefinita corrisponde all'eventuale etichetta di variabile o nome di variabile della variabile sorgente con l'aggiunta di *(Segmentata)* alla fine dell'etichetta.

Minimo e Massimo. I valori minimo e massimo della variabile attualmente selezionata, in base ai casi esaminati ed esclusi i valori definiti come valori mancanti definiti dall'utente.

Valori non mancanti. L'istogramma mostra la distribuzione dei valori non mancanti della variabile attualmente selezionata, in base ai casi esaminati.

- Dopo aver definito le sezioni per la nuova variabile, nell'istogramma vengono visualizzate linee verticali per indicare i punti di divisione che delimitano le sezioni.
- È possibile trascinare le linee dei punti di divisione in posizioni diverse sull'istogramma, modificando gli intervalli delle sezioni.
- È possibile rimuovere le sezioni trascinando le linee dei punti di divisione all'esterno dell'istogramma.

Nota: l'istogramma (in cui sono visualizzati valori non mancanti) e i valori minimo e massimo sono basati sui valori esaminati. Se nell'analisi non vengono inclusi tutti i casi, è possibile che la distribuzione effettiva non venga rappresentata in modo preciso, soprattutto se il file di dati è stato ordinato in base alla variabile selezionata. Se non viene esaminato alcun caso, non sarà disponibile alcuna informazione sulla distribuzione dei valori.

Griglia. Visualizza i valori che definiscono i punti finali superiori e le etichette di valore opzionali di ogni sezione.

- **Valore.** Il valore che definisce i punti finali superiori di ogni sezione. È possibile specificare valori o utilizzare l'opzione Crea punti divisione per creare automaticamente sezioni basate sui criteri selezionati. Per default, viene automaticamente incluso un punto di divisione con il valore *MAX*. Questa sezione contiene tutti i valori non mancanti superiori agli altri punti di divisione. La sezione delimitata dal punto di divisione inferiore include tutti i valori non mancanti inferiori o uguali a tale valore (o solo inferiori al valore, a seconda delle modalità di definizione dei punti finali superiori).
- **Etichetta.** Opzionale, etichette descrittive per i valori della nuova variabile segmentata. Dal momento che i valori della nuova variabile sono semplicemente interi sequenziali da 1 a *n*, le etichette descrittive che indicano ciò che il valore rappresenta possono rivelarsi molto utili. È possibile specificare le etichette o utilizzare l'opzione Crea etichette per crearle automaticamente.

Per eliminare una sezione dalla griglia

▶ Fare clic con il pulsante destro del mouse sulla cella *Valore* o *Etichetta* relativa alla sezione.

▶ Dal menu di scelta rapida, scegliere Elimina la riga.

Nota: se si elimina la sezione *MAX*, a tutti i casi con valori superiori all'ultimo valore di punto di divisione specificato verrà assegnato il valore mancante di sistema della nuova variabile.

Per eliminare tutte le etichette o tutte le selezioni definite

▶ Fare clic con il pulsante destro del mouse in un punto qualsiasi della griglia.

▶ Dal menu di scelta rapida, scegliere Elimina tutte le etichette o Elimina tutti i punti di divisione.

Punti finali superiori. Controlla il trattamento dei valori finali superiori nella colonna *Valore* della griglia.

- **Inclusi (<=).** I casi con il valore specificato nella cella *Valore* sono inclusi nella categoria segmentata. Ad esempio, se si specificano come valori 25, 50 e 75, i casi con il valore 25 verranno inclusi nella prima sezione, in quanto tale sezione conterrà tutti i casi con valori inferiori o uguali a 25.
- **Esclusi (<).** I casi con il valore specificato nella cella *Valore* sono esclusi dalla categoria segmentata e invece inclusi nella sezione successiva. Ad esempio, se si specificano come valori 25, 50 e 75, i casi con il valore 25 verranno inclusi nella seconda sezione anziché nella prima, in quanto la prima sezione conterrà tutti i casi con valori inferiori a 25.

Crea punti divisione. Genera automaticamente categorie segmentate per intervalli di larghezza uguale, intervalli con lo stesso numero di casi o intervalli basati su deviazioni standard. Non è disponibile se non viene esaminato alcun caso. Per ulteriori informazioni, vedere "Generazione automatica di categorie segmentate" a pag. 155.

Crea etichette. Genera etichette descrittive per i valori interi sequenziali della nuova variabile segmentata, in base ai valori nella griglia e al trattamento specifico dei punti finali superiori (inclusi o esclusi).

Scala inversa. Per default, i valori della nuova variabile segmentata sono interi sequenziali crescenti da 1 a *n*. L'inversione della scala modifica tali valori in interi sequenziali decrescenti da *n* a 1.

Copia sezioni. È possibile copiare le specifiche di segmentazione di un'altra variabile nella variabile attualmente selezionata o dalla variabile selezionata in più variabili diverse. Per ulteriori informazioni, vedere "Copia di categorie segmentate" a pag. 158.

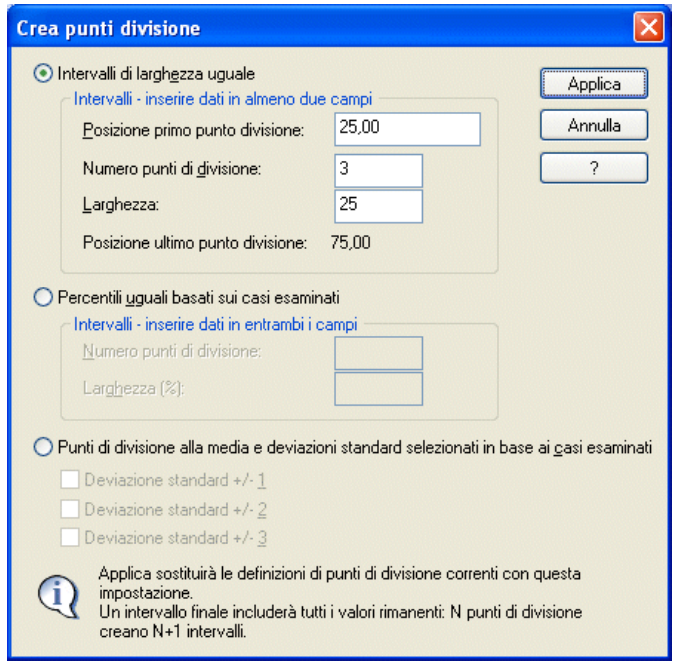
Generazione automatica di categorie segmentate

La finestra di dialogo Crea punti divisione consente di generare automaticamente categorie segmentate in base a criteri selezionati.

Per utilizzare la finestra di dialogo Crea punti divisione

- Selezionare una variabile dall'Elenco variabili esaminate.
- Fare clic su Crea punti divisione.
- Selezionare i criteri per la generazione dei punti di divisione che definiranno le categorie segmentate.
- Fare clic su Applica.

Figura 7-15
Finestra di dialogo Crea punti divisione


Crea punti divisione
Intervalli di larghezza uguale
Intervalli - inserire dati in almeno due campi
Posizione primo punto divisione:
25,00
Numero punti di divisione:
3
Larghezza:
25
Posizione ultimo punto divisione:
75,00
Applica
Annulla
?
Percentili uguali basati sui casi esaminati
Intervalli - inserire dati in entrambi i campi
Numero punti di divisione:
Larghezza (%):
Punti di divisione alla media e deviazioni standard selezionati in base ai casi esaminati
Deviazione standard +/- 1
Deviazione standard +/- 2
Deviazione standard +/- 3
Applica sostituirà le definizioni di punti di divisione correnti con questa impostazione.
Un intervallo finale includerà tutti i valori rimanenti: N punti di divisione creano N+1 intervalli.

Nota: la finestra di dialogo Crea punti divisione non è disponibile se non viene esaminato alcun caso.

Intervalli di larghezza uguale. Genera categorie segmentate di larghezza uguale (ad esempio, 1–10, 11–20, 21–30 e così via) in base a due dei tre criteri seguenti:

- **Posiz. primo punto div.** Valore che definisce il punto superiore della categoria segmentata inferiore (ad esempio, il valore 10 indica un intervallo che include tutti i valori fino a 10).
- **N. punti di divisione.** Il numero di categorie segmentate corrisponde al numero di punti di divisione *più uno*. Ad esempio, 9 punti di divisione generano 10 categorie segmentate.
- **Larghezza.** Larghezza di ogni intervallo. Ad esempio, il valore 10 determina la segmentazione di *Età in anni* in intervalli di 10 anni.

Percentili uguali basati sui casi esaminati. Genera categorie segmentate con un numero uguale di casi in ogni sezione (mediante l'algoritmo empirico per i percentili), in base a uno dei criteri seguenti:

- **N. punti di divisione.** Il numero di categorie segmentate corrisponde al numero di punti di divisione *più uno*. Ad esempio, tre punti di divisione generano quattro sezioni percentili (quartili) contenenti ognuna il 25% dei casi.
- **Larghezza (%).** Larghezza di ogni intervallo, espressa come percentuale del numero totale di casi. Ad esempio, il valore 33,3 determina la creazione di tre categorie segmentate (due punti di divisione) contenenti ognuna il 33,3% dei casi.

Se la variabile sorgente contiene un numero relativamente limitato di valori distinti o un numero di casi con lo stesso valore, è possibile che venga creato un numero di sezioni inferiore a quello richiesto. Se esistono più valori identici per un punto di divisione, questi vengono inseriti tutti nello stesso intervallo; quindi è possibile che le percentuali effettive non siano sempre esattamente uguali.

Punti di divisione alla media e deviaz. stand. selezionati in base ai casi esaminati. Genera categorie segmentate basate sui valori di media e deviazione standard della distribuzione della variabile.

- Se non si seleziona nessuno degli intervalli di deviazione standard, verranno create due categorie segmentate che utilizzano la media come punto di divisione.
- È possibile selezionare qualsiasi combinazione di intervalli di deviazione standard basata su una, due e/o tre deviazioni standard. Ad esempio, la selezione di tutte e tre determina la creazione di otto categorie segmentate, ovvero sei sezioni negli intervalli di una deviazione standard e due sezioni per i casi superiori alle tre deviazioni sopra e sotto la media.

In una distribuzione normale, il 68% dei casi è compreso in un intervallo pari a una deviazione standard della media, il 95% in un intervallo pari a due deviazioni standard e il 99% in un intervallo pari a tre deviazioni standard. La creazione di categorie segmentate basate su deviazioni standard può determinare la creazione di sezioni definite esterne all'intervallo di dati effettivo o anche esterne all'intervallo di valori di dati possibili (ad esempio, in un intervallo di stipendi negativo).

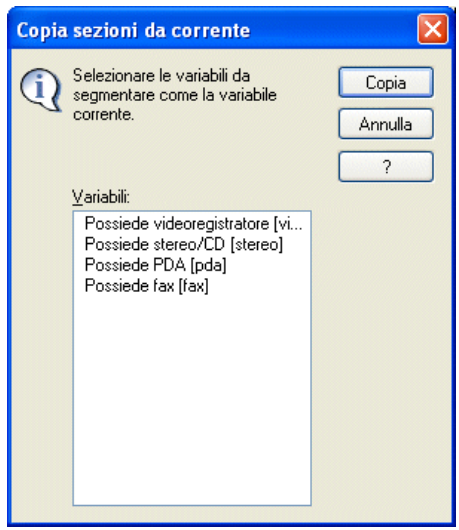
Nota: i calcoli dei percentili e delle deviazioni standard sono basati sui casi esaminati. Se si limita il numero di casi esaminati, è possibile che le sezioni risultanti non contengano la proporzione di casi desiderata in tali sezioni, soprattutto se il file di dati è ordinato in base alla variabile sorgente. Ad esempio, se si limita l'analisi ai primi

100 casi di un file di dati con 1000 casi e il file di dati viene ordinato in base all'ordine ascendente di età dei rispondenti, anziché sezioni percentili di età contenenti ognuna il 25% dei casi è possibile che le prime tre sezioni create contengano ognuna circa il 3,3% dei casi e l'ultima sezione contenga il 90% dei casi.

Copia di categorie segmentate

Quando si creano categorie segmentate per una o più variabili, è possibile copiare le specifiche di segmentazione di un'altra variabile nella variabile attualmente selezionata o dalla variabile selezionata in più variabili diverse.

Figura 7-16
Copia di sezioni nella variabile corrente o dalla variabile corrente



Per copiare specifiche di segmentazione

- Definire categorie segmentate per almeno una variabile ma *non* fare clic su OK o su Incolla.
- Selezionare una variabile dall'Elenco variabili esaminate per cui sono state definite categorie segmentate.
- Fare clic su Ad altre variabili.

- Selezionare le variabili per cui si desidera creare nuove variabili con le stesse categorie segmentate.
- Fare clic su Copia.

oppure

- Selezionare una variabile dall'Elenco variabili esaminate in cui si desidera copiare categorie segmentate definite.
- Fare clic su Da un'altra variabile.
- Selezionare la variabile con le categorie segmentate definite che si desidera copiare.
- Fare clic su Copia.

Se per la variabile da cui si copiano le specifiche di segmentazione sono state specificare etichette di valori, verranno copiate anche tali etichette.

Nota: dopo aver fatto clic su OK nella finestra principale di Segmentazione grafica per creare nuove variabili segmentate (o dopo aver chiuso tale finestra in altro modo), non sarà possibile utilizzare Segmentazione grafica per copiare le categorie segmentate in altre variabili.

Valori mancanti definiti dall'utente in Segmentazione grafica

I valori definiti come valori mancanti definiti dall'utente (ovvero i valori identificati come codici per dati mancanti) per la variabile sorgente non sono inclusi nelle categorie segmentate per la nuova variabile. I valori mancanti definiti dall'utente per le variabili sorgente vengono copiati come valori mancanti definiti dall'utente della nuova variabile insieme a tutte le etichette di valori definiti per i codici di valori mancanti.

Se un codice di valore mancante è in conflitto con uno dei valori delle categorie segmentate per la nuova variabile, il conflitto viene risolto utilizzando per la registrazione di tale codice nella nuova variabile un valore corrispondente all'aggiunta di 100 al valore più alto delle categorie segmentate. Ad esempio, se il valore 1 è definito come valore mancante definito dall'utente per la variabile sorgente e la nuova variabile contiene sei categorie segmentate, tutti i casi con valore 1 avranno valore 106 per la nuova variabile e 106 verrà definito come valore mancante definito dall'utente. Se il valore mancante definito dall'utente per la variabile sorgente include

un'etichetta di valore definita, tale etichetta verrà mantenuta come etichetta del valore registrato per la nuova variabile.

Nota: se la variabile sorgente include un intervallo definito di valori mancanti definiti dall'utente nel formato *LO-n*, dove *n* corrisponde a un numero positivo, i valori mancanti definiti dall'utente corrispondenti per la nuova variabile saranno numeri negativi.

Capitolo

8

Trasformazioni di dati

Nella situazione ideale, i dati sono perfettamente adatti al tipo di analisi che si desidera eseguire e le relazioni tra le variabili sono lineari oppure nettamente ortogonali. Purtroppo questo caso non si verifica quasi mai. Le analisi preliminari possono rivelare la presenza di schemi di codifica difficili o di errori di codifica, oppure può essere necessario eseguire trasformazioni di dati per identificare la relazione effettiva tra le variabili.

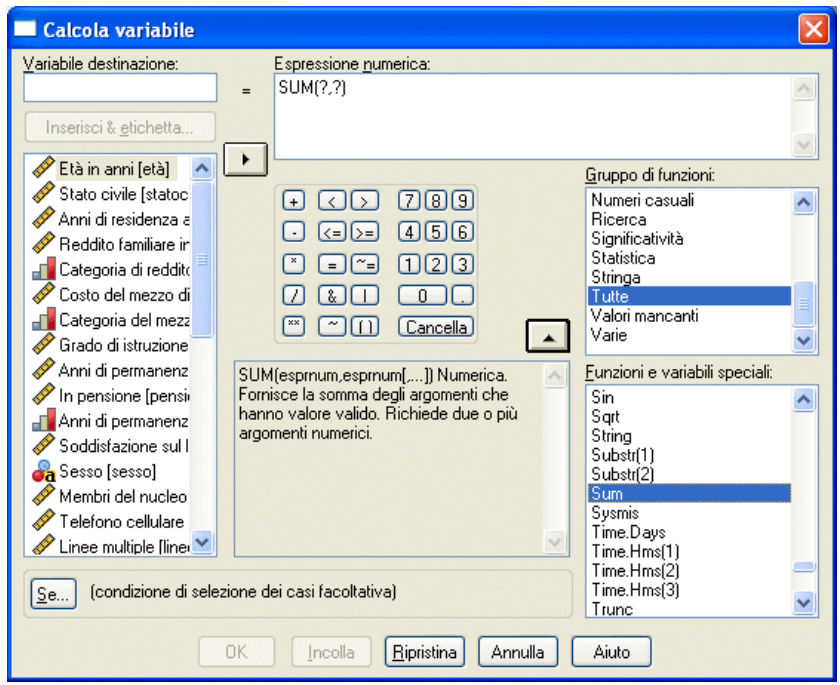
È possibile eseguire trasformazioni di dati rappresentate da attività semplici, ad esempio la compressione delle categorie per l'analisi, o da attività più complesse, come la creazione di nuove variabili in base ad equazioni complesse e istruzioni condizionali.

Calcolo delle variabili

Utilizzare la finestra di dialogo Calcola per calcolare i valori di una variabile basata su trasformazioni numeriche di altre variabili.

- È possibile calcolare i valori di variabili numeriche o variabili stringa (alfanumeriche).
- È possibile creare nuove variabili oppure sostituire i valori di variabili esistenti. Per le nuove variabili è inoltre possibile specificare il tipo e l'etichetta.
- È possibile calcolare i valori selettivamente per sottoinsiemi di dati in base a condizioni logiche.
- È possibile utilizzare più di 70 funzioni predefinite, ad esempio funzioni aritmetiche, statistiche, di distribuzione e stringa.

Figura 8-1
Finestra di dialogo Calcola variabile



Per calcolare variabili

- Dai menu, scegliere:

  Trasforma
  Calcola...

- Digitare il nome di una sola variabile di destinazione. È possibile aggiungere al file di dati attivo una variabile nuova o esistente.

- Per creare un'espressione, è possibile incollare o digitare direttamente i componenti nel campo Espressione.

  - È possibile incollare le funzioni o le variabili di sistema normalmente utilizzate selezionando un gruppo nell'elenco delle funzioni e facendo doppio clic su una funzione o variabile nell'elenco delle funzioni e variabili speciali. È possibile anche selezionare la funzione o la variabile, quindi fare clic sulla freccia vicino all'elenco delle funzioni. Immettere i valori per tutti i parametri contrassegnati

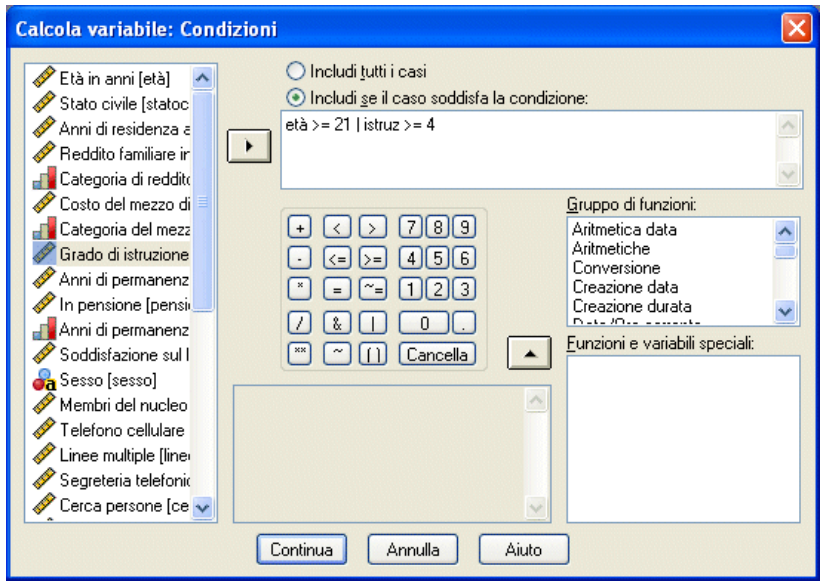
con punti interrogativi (solo per le funzioni). Il gruppo di funzioni denominato Tutti contiene un elenco di tutte le funzioni e variabili di sistema disponibili. L'area dedicata alla finestra dialogo visualizza una breve descrizione della funzione o variabile correntemente selezionata.

- Le costanti stringa devono essere incluse tra virgolette o apostrofi.
- Se i valori contengono decimali, usare il punto (.) come separatore decimale.
- Per le nuove variabili stringa, è necessario selezionare anche Tipo ed etichetta per specificare il tipo di dati.

Calcola variabile: Condizioni

Nella finestra di dialogo Condizioni è possibile applicare trasformazioni di dati al sottoinsieme di casi selezionato utilizzando le espressioni logiche. Un'espressione logica restituisce un valore *vero*, *falso*, o *mancante* per ciascun caso.

Figura 8-2
Finestra di dialogo Calcola variabile: Condizioni



- Se il risultato di un'espressione logica è *vero*, il caso verrà incluso nel sottoinsieme selezionato.

- Se il risultato di un'espressione logica è *falso* o *mancante*, il caso non viene incluso nel sottoinsieme selezionato.
- La maggior parte delle espressioni condizionali utilizza uno o più dei sei operatori relazionali (<, >, <=, >=, = e ~=) della calcolatrice.
- Le espressioni logiche possono includere nomi di variabili, costanti, operatori aritmetici, funzioni numeriche e di altro tipo, variabili logiche e operatori relazionali.

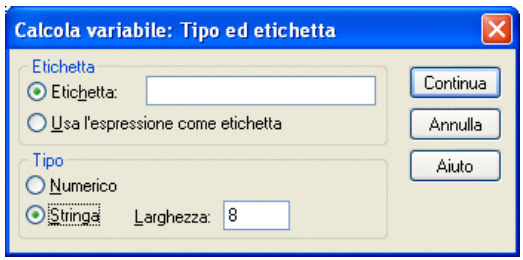
Calcola variabile: Tipo ed etichetta

Per impostazione predefinita, le nuove variabili calcolate sono numeriche. Per calcolare una nuova variabile stringa, è necessario specificare il tipo di dati e la larghezza.

Etichetta. Etichetta di variabile descrittiva e opzionale, con una lunghezza massima di 120 caratteri. È possibile immettere un'etichetta oppure utilizzare i primi 110 caratteri dell'espressione Calcola.

Tipo. Le variabili calcolate possono essere di tipo numerico o stringa (alfanumeriche). Nei calcoli non è possibile utilizzare le variabili stringa.

Figura 8-3
Finestra di dialogo Tipo ed etichetta.



Funzioni

Sono supportati molti tipi di funzioni, inclusi:

- Funzioni aritmetiche
- Funzioni statistiche
- Funzioni stringa

- Funzioni di data e di ora
- Funzioni di distribuzione
- Funzioni per variabili casuali
- Funzioni per valori mancanti
- Funzioni di punteggio (solo per SPSS Server)

Ricercare la voce funzioni nell'indice dell'Aiuto in linea per l'elenco completo delle funzioni disponibili.

Valori mancanti nelle funzioni

Nelle funzioni e nelle espressioni matematiche semplici i valori mancanti vengono trattati in modo diverso. Nell'espressione seguente:

```
(var1+var2+var3)/3
```

il risultato è mancante se un caso include un valore mancante per una delle tre variabili.

Nell'espressione seguente:

```
MEAN(var1, var2, var3)
```

il risultato è mancante solo se il caso include valori mancanti per tutte e tre le variabili.

Per quanto riguarda le funzioni statistiche, è possibile specificare il numero minimo di argomenti che devono includere valori non mancanti. Per eseguire questa operazione, digitare un punto e il numero minimo dopo il nome della funzione, ad esempio:

```
MEAN.2(var1, var2, var3)
```

Generatore di numeri casuale

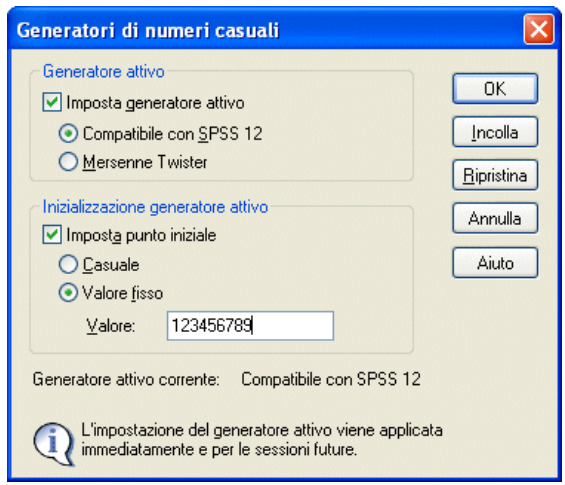
La finestra di dialogo Generatore di numeri casuale permette di selezionare il generatore di numeri casuale e di impostare il valore iniziale della sequenza da utilizzare per riprodurre una sequenza di numeri casuali.

Generatore attivo. Sono disponibili due diversi generatori di numeri casuali:

- **Compatibile con SPSS 12.** Il generatore di numeri casuali usato in SPSS 12 e versioni precedenti. Usare questo generatore se è necessario riprodurre i risultati casuali ottenuti da versioni precedenti di SPSS.
- **Mersenne Twister.** Un nuovo generatore di numeri casuali più affidabile per scopi di simulazione. Usare questo generatore se non serve riprodurre i risultati casuali ottenuti da versioni precedenti di SPSS.

Inizializzazione generatore attivo. Il seme dei numeri casuali cambia ogni volta che viene generato un numero casuale da utilizzare per le trasformazioni (ad esempio le funzioni di distribuzione casuale), il campionamento casuale e la ponderazione dei casi. Per replicare una sequenza di numeri casuali, impostare il valore del punto di inizializzazione iniziale prima di eseguire l'analisi basata su numeri casuali. Il valore deve essere un numero positivo intero.

Figura 8-4
Finestra di dialogo Generatori di numeri casuali



Per selezionare il generatore di numeri casuale e/o impostare il valore di inizializzazione

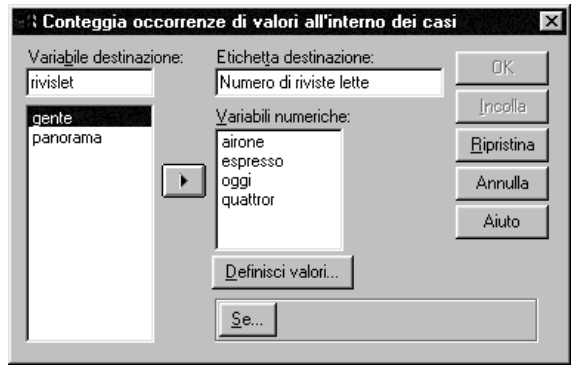
▶ Dai menu, scegliere:

Trasforma
  Generatore di numeri casuale

Conteggia occorrenze di valori all'interno dei casi

Questa finestra di dialogo consente di creare una variabile per il conteggio delle occorrenze dello stesso valore in una lista di variabili relative allo stesso caso. Ad esempio, un'indagine può contenere una lista di riviste con le caselle *sì*/*no* in cui è possibile indicare quali riviste vengono lette dai rispondenti. È possibile specificare il numero di risposte *affermative* fornite da ciascuna persona per creare una nuova variabile che contenga il numero totale di riviste lette.

Figura 8-5
Finestra di dialogo Conteggia occorrenze di valori all'interno dei casi



Per contare le occorrenze dei valori all'interno dei casi

- Dai menu, scegliere:
  Trasforma
  Conteggia...

- Specificare il nome della variabile di destinazione.

- Selezionare due o più variabili dello stesso tipo (numerico o stringa).

- Fare clic su Definisci valori e specificare i valori da contare.

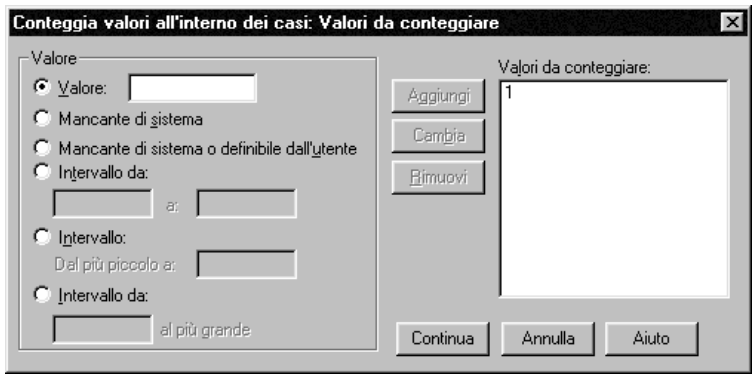
  È inoltre possibile definire un sottoinsieme di casi in cui contare le occorrenze dei valori.

Conteggia valori all'interno dei casi: Valori da conteggiare

Il valore della variabile di destinazione (nella finestra di dialogo principale) aumenta di 1 ogni volta che le variabili selezionate corrispondono a una specificazione della lista Valori da conteggiare. Se un caso corrisponde a più specificazioni per la stessa variabile, la variabile di destinazione aumenterà più volte per tale variabile.

Le specificazioni dei valori possono includere valori singoli, valori mancanti o valori mancanti di sistema e intervalli. Gli intervalli includono i relativi punti finali e i valori mancanti definiti dall'utente compresi in ciascun intervallo.

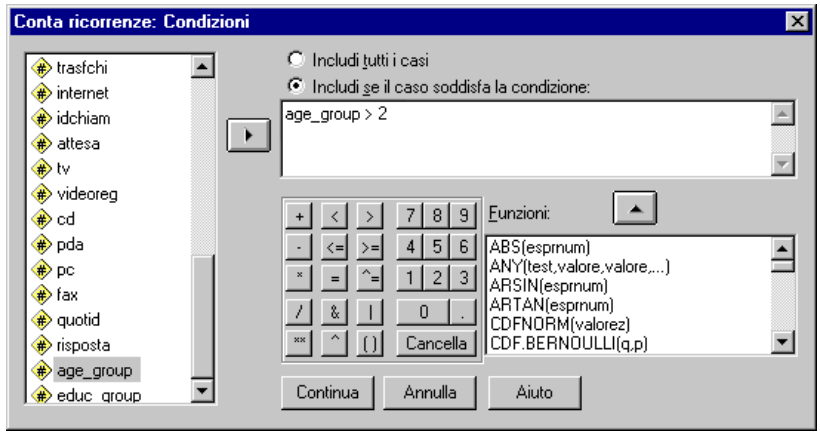
Figura 8-6
Finestra di dialogo Valori da conteggiare



Conta ricorrenze: Condizioni

La finestra di dialogo Condizioni permette di contare le ricorrenze dei valori di un sottoinsieme specifico di casi mediante espressioni condizionali. Un'espressione logica restituisce un valore *vero*, *falso*, o *mancante* per ciascun caso.

Figura 8-7
Finestra di dialogo Conta ricorrenze: Condizioni



Per informazioni generali sull'uso della finestra di dialogo Condizioni, vedere "Calcola variabile: Condizioni" a pag. 163.

Ricodifica dei valori

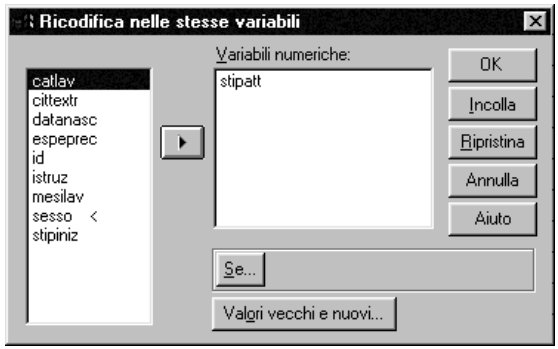
Per modificare i valori, è possibile ricodificarli. Questo risulta particolarmente utile per comprimere o combinare più categorie. È possibile ricodificare i valori all'interno delle variabili esistenti oppure creare nuove variabili in base ai valori ricodificati delle variabili esistenti.

Ricodifica nelle stesse variabili

La finestra di dialogo Ricodifica nelle stesse variabili permette di riassegnare i valori delle variabili esistenti o di comprimere gli intervalli dei valori esistenti in nuovi valori. È possibile, ad esempio, comprimere i salari in categorie di intervalli di salari.

È possibile ricodificare variabili numeriche e variabili stringa. Se si selezionano più variabili, queste devono essere dello stesso tipo. Non è possibile ricodificare contemporaneamente variabili stringa e numeriche.

Figura 8-8
Finestra di dialogo Ricodifica nelle stesse variabili



Per ricodificare i valori di una variabile

- Dai menu, scegliere:

  Trasforma
  Ricodifica
  Nelle stesse variabili...

- Selezionare le variabili che si desidera ricodificare. Se vengono selezionate più variabili, devono essere dello stesso tipo (numerico o stringa).

- Fare clic su Valori vecchi e nuovi e specificare le modalità di ricodifica dei valori.

  È inoltre possibile definire un sottoinsieme di casi da ricodificare. La finestra di dialogo Condizioni è la stessa di quella descritta per Conta occorrenze.

Ricodifica nelle stesse variabili: Valori vecchi e nuovi

In questa finestra di dialogo è possibile definire i valori da ricodificare. Tutte le specificazioni dei valori devono essere dello stesso tipo di dati (numerico o stringa) delle variabili selezionate nella finestra di dialogo principale.

Vecchio valore. I valori da ricodificare. È possibile ricodificare singoli valori, intervalli di valori o valori mancanti. I valori e gli intervalli mancanti di sistema non possono essere selezionati come variabili stringa in quanto nessuno dei due concetti è valido per tali variabili. Gli intervalli includono i relativi punti finali e i valori mancanti definiti dall'utente compresi in ciascun intervallo.

- **Valore (Ricodifica).** Valore discreto da ricodificare in un nuovo valore. Il valore deve essere dello stesso tipo (numerico o stringa) delle variabili da ricodificare.
- **Mancanti di sistema.** Il valore mancante di sistema rappresenta dati numerici non noti. Viene usato come risultato di calcoli che non possono essere eseguiti ed è assegnato quando campi completamente vuoti vengono letti come dati numerici. L'editor dei dati e molti comandi che visualizzano i singoli valori rappresentano il valore mancante di sistema tramite un punto. Per le stringhe non può esserci valore mancante di sistema, poiché in una variabile stringa qualsiasi carattere è valido.
- **Valori mancanti di sistema o definibili dall'utente.** Osservazioni con valori che sono stati dichiarati mancanti dall'utente o a cui è stato assegnato il valore mancante di sistema, indicato da un punto (.).
- **Intervallo di valori mancanti (Definisci valori mancanti).** Un intervallo di valori, inclusi gli estremi specificati. Non disponibile per variabili stringa.
- **Tutti gli altri valori.** Tuti i valori non esplicitamente inseriti nell'elenco dei vecchi valori. Aggiunge una clausola ELSE all'elenco delle ricodifiche.

Nuovo valore. Il singolo valore in cui viene ricodificato ciascun vecchio valore o intervallo di valori. È possibile specificare un valore oppure assegnare il valore mancante di sistema.

- **Valore.** Valore destinato a sostituire uno o più vecchi valori. Deve essere dello stesso tipo (numerico o stringa) del vecchio valore corrispondente.
- **Mancanti di sistema.** Sostituisce al vecchio valore specificato il valore mancante di sistema. I valori mancanti vengono esclusi dalla maggior parte delle analisi. Non è applicabile a variabili di tipo stringa.

Vecchio->Nuovo. La lista delle specificazioni che verranno utilizzate per ricodificare le variabili. È possibile aggiungere, modificare e rimuovere le specifiche dall'elenco. L'elenco verrà ordinato automaticamente in base alla specifiche del vecchio valore e con il seguente ordine: valori unici, valori mancanti, intervalli e tutti gli altri valori. Se nell'elenco viene modificata una specifica di ricodifica, la procedura riordina automaticamente l'elenco, se necessario, per mantenere tale ordine.

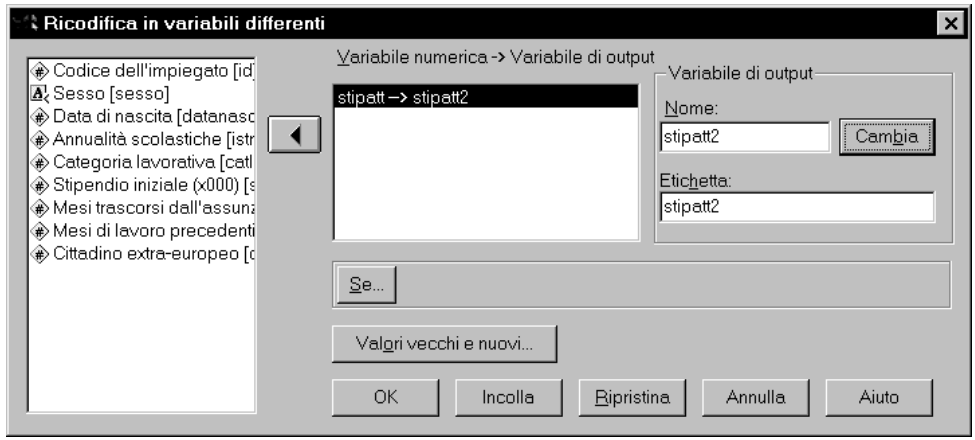
Figura 8-9
Finestra di dialogo Valori vecchi e nuovi

Ricodifica in variabili differenti

La finestra di dialogo Ricodifica in variabili differenti permette di riassegnare i valori delle variabili esistenti o di comprimere gli intervalli dei valori esistenti nei nuovi valori della variabile. È possibile, ad esempio, comprimere i salari in una nuova variabile che contenga categorie di intervalli di salari.

- È possibile ricodificare variabili numeriche e variabili stringa.
- È possibile ricodificare variabili numeriche in variabili stringa e viceversa.
- Se si selezionano più variabili, queste devono essere dello stesso tipo. Non è possibile ricodificare contemporaneamente variabili stringa e numeriche.

Figura 8-10
Finestra di dialogo Ricodifica in variabili differenti



Per ricodificare i valori di una variabile in una nuova variabile

- Dai menu, scegliere:

  Trasforma
  Ricodifica
  In variabili differenti...

- Selezionare le variabili che si desidera ricodificare. Se vengono selezionate più variabili, devono essere dello stesso tipo (numerico o stringa).

- Specificare un (nuovo) nome di variabile di output per ciascuna nuova variabile e fare clic su Cambia.

- Fare clic su Valori vecchi e nuovi e specificare le modalità di ricodifica dei valori.

  È inoltre possibile definire un sottoinsieme di casi da ricodificare. La finestra di dialogo Condizioni è la stessa di quella descritta per Conta occorrenze.

Ricodifica in variabili differenti: Valori vecchi e nuovi

In questa finestra di dialogo è possibile definire i valori da ricodificare.

Vecchio valore. I valori da ricodificare. È possibile ricodificare singoli valori, intervalli di valori o valori mancanti. I valori e gli intervalli mancanti di sistema non possono essere selezionati come variabili stringa in quanto nessuno dei due concetti è

valido per tali variabili. I vecchi valori devono avere lo stesso tipo di dati (numerico o stringa) della variabile originale. Gli intervalli includono i relativi punti finali e i valori mancanti definiti dall'utente compresi in ciascun intervallo.

- **Valore (Ricodifica).** Valore discreto da ricodificare in un nuovo valore. Il valore deve essere dello stesso tipo (numerico o stringa) delle variabili da ricodificare.
- **Mancanti di sistema.** Il valore mancante di sistema rappresenta dati numerici non noti. Viene usato come risultato di calcoli che non possono essere eseguiti ed è assegnato quando campi completamente vuoti vengono letti come dati numerici. L'editor dei dati e molti comandi che visualizzano i singoli valori rappresentano il valore mancante di sistema tramite un punto. Per le stringhe non può esserci valore mancante di sistema, poiché in una variabile stringa qualsiasi carattere è valido.
- **Valori mancanti di sistema o definibili dall'utente.** Osservazioni con valori che sono stati dichiarati mancanti dall'utente o a cui è stato assegnato il valore mancante di sistema, indicato da un punto (.).
- **Intervallo di valori mancanti (Definisci valori mancanti).** Un intervallo di valori, inclusi gli estremi specificati. Non disponibile per variabili stringa.
- **Tutti gli altri valori.** Tuti i valori non esplicitamente inseriti nell'elenco dei vecchi valori. Aggiunge una clausola ELSE all'elenco delle ricodifiche.

Nuovo valore. Il singolo valore in cui viene ricodificato ciascun vecchio valore o intervallo di valori. I nuovi valori possono essere di tipo numerico o stringa.

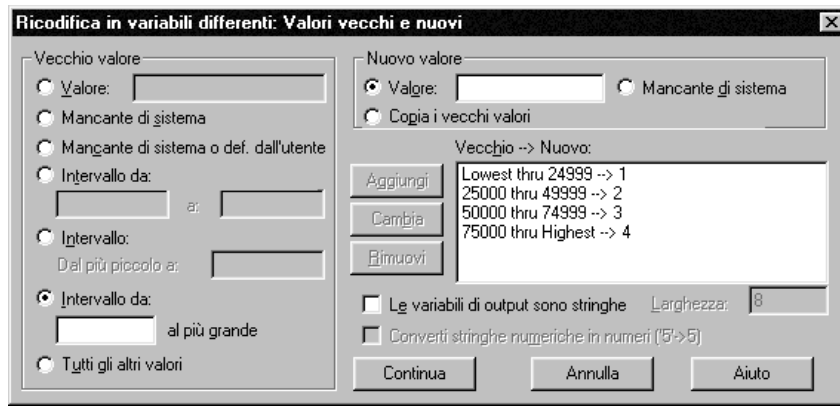
- **Valore.** Valore destinato a sostituire uno o più vecchi valori. Deve essere dello stesso tipo (numerico o stringa) del vecchio valore corrispondente.
- **Mancanti di sistema.** Sostituisce al vecchio valore specificato il valore mancante di sistema. I valori mancanti vengono esclusi dalla maggior parte delle analisi. Non è applicabile a variabili di tipo stringa.
- **Copia i vecchi valori.** Mantiene i vecchi valori, quando non è richiesta la loro ricodifica. Attenzione a specificare esplicitamente i valori da non ricodificare. Per default, infatti, la nuova variabile assumerà il valore mancante di sistema quando la variabile originale assume valori non esplicitamente inclusi nell'elenco di quelli da ricodificare.

Le variabili di output sono stringhe (Ricodifica). Definisce le nuove variabili ricodificate come stringhe. Le variabili originali possono essere numeriche o stringhe.

Converti stringhe numeriche in numeri. Converte in numeri le stringhe contenenti valori numerici (le cifre da 0 a 9 e, opzionalmente, il segno). I casi con valori non numerici assumeranno valore mancante di sistema nella variabile di destinazione.

Vecchio->Nuovo. La lista delle specificazioni che verranno utilizzate per ricodificare le variabili. È possibile aggiungere, modificare e rimuovere le specifiche dall'elenco. L'elenco verrà ordinato automaticamente in base alla specifiche del vecchio valore e con il seguente ordine: valori unici, valori mancanti, intervalli e tutti gli altri valori. Se nell'elenco viene modificata una specifica di ricodifica, la procedura riordina automaticamente l'elenco, se necessario, per mantenere tale ordine.

Figura 8-11
Finestra di dialogo Valori vecchi e nuovi



Classifica casi

La finestra di dialogo Classifica casi permette di creare nuove variabili contenenti ranghi, punteggi normali ed esponenziali e valori percentili per le variabili numeriche.

I nomi delle nuove variabili e le etichette delle variabili descrittive verranno generate automaticamente in base al nome originale della variabile e alle misure selezionate. Nella tabella riassuntiva verranno elencate le variabili originali, le nuove variabili e le etichette delle variabili.

Se lo si desidera, è possibile:

- Classificare i casi in ordine crescente o decrescente.
- Organizzare ranghi in sottogruppi selezionando una o più variabili di raggruppamento per la lista Per. Verranno creati ranghi all'interno di ciascun gruppo. I gruppi vengono definiti dalla combinazione dei valori delle variabili di raggruppamento. Ad esempio, se si seleziona *sesso* e *minoranza* come variabili di raggruppamento, verranno creati ranghi per ciascuna combinazione di *sesso* e *minoranza*.

Figura 8-12
Finestra di dialogo Classifica casi

Per classificare i casi

▶ Dai menu, scegliere:

Trasforma
  Classifica casi

▶ Selezionare una o più variabili da classificare. È possibile classificare solo variabili numeriche.

È possibile classificare i casi in ordine crescente o decrescente e organizzarli in sottogruppi.

Classifica casi: Tipi

È possibile selezionare più metodi di classifica. Per ciascun metodo verrà creata una variabile di rango distinta. Sono disponibili diversi metodi di classifica, ovvero rango semplice, punteggi esponenziali, punteggi frazionari e percentili. È inoltre possibile creare classifiche in base alle stime di proporzione e ai punteggi normali.

Rango. Il rango semplice, espresso come intero positivo. Il valore della nuova variabile corrisponde al rango.

Punteggio esponenziale. La nuova variabile contiene punteggi esponenziali basati su una distribuzione esponenziale.

Rango frazionario. Il valore della nuova variabile è uguale al rango diviso la somma dei pesi dei casi non mancanti.

Rango frazionario come percentuale. Ogni rango è diviso per il numero di casi con valori validi e moltiplicato per 100.

Somma dei pesi di caso. Il valore della nuova variabile è uguale alla somma dei pesi di caso. La nuova variabile è una costante per tutti i casi nello stesso gruppo.

N percentili. I ranghi sono definiti in base a gruppi percentili, dove ogni gruppo contiene approssimativamente lo stesso numero di casi. Per esempio, 4 percentili assegnerebbe rango 1 ai casi compresi entro il 25° percentile, 2 ai casi compresi fra il 25° e il 50° percentile, 3 ai casi compresi fra il 50° e il 75° percentile, 4 a quelli oltre il 75°.

Stime di proporzione. Sono stime della proporzione cumulata di distribuzione in corrispondenza di un rango.

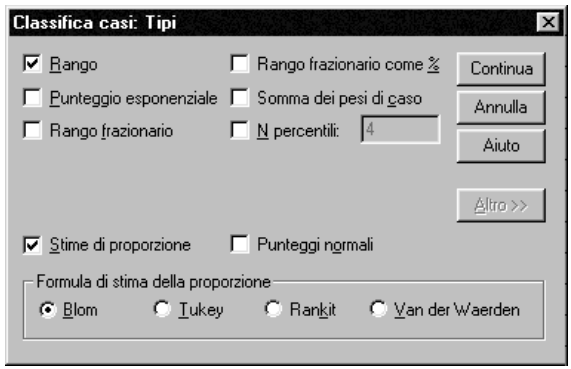
Punteggi normali. I punteggi normali sono i punteggi Z corrispondenti alla proporzione cumulata stimata.

Formula di stima della proporzione. Per le stime di proporzione e i punteggi normali è possibile selezionare la formula di stima della proporzione: Blom, Tukey, Rankit o Van der Waerden.

- **Blom.** Crea una nuova classifica basata su stime di proporzione. Usa la formula (r-3/8) / (w+1/4), dove r è il rango e w è la somma dei pesi dei casi.
- **Tukey.** Usa la formula (r-1/3) / (p+1/3), dove r è il rango e p è la somma dei pesi.

- **Rankit.** Usa la formula (r-1/2) / n, dove n è il numero di osservazioni e r è il rango, che varia fra 1 e n.
- **Van der Waerden.** Trasformazione di Van der Waerden, definita dalla la formula r/(p+1), dove p è la somma dei pesi e r è il rango, che varia da 1 a p.

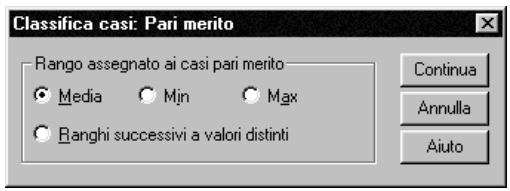
Figura 8-13
Finestra di dialogo Classifica casi: Tipi



Classifica casi: Pari merito

Questa finestra di dialogo consente di controllare il metodo di assegnazione dei ranghi ai casi con lo stesso valore della variabile originale.

Figura 8-14
Finestra di dialogo Classifica casi: Pari merito



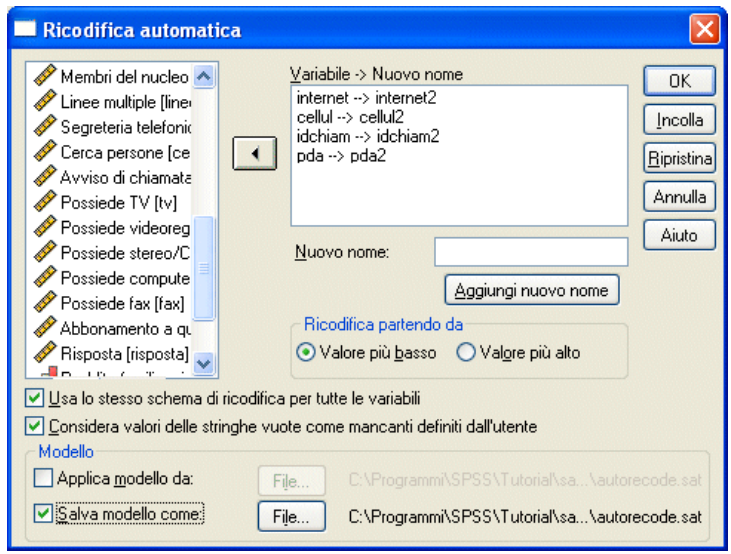
Nella tabella che segue vengono indicati i diversi metodi di assegnazione dei ranghi ai valori di pari merito:

Valore	Media	Basso	Alto	Sequenziale
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Ricodifica automatica

La finestra di dialogo Ricodifica automatica permette di convertire i valori stringa e numerici in numeri interi consecutivi. Se i codici di categoria non sono sequenziali, le celle vuote risultanti comportano prestazioni inferiori e maggiori requisiti di memoria per l'esecuzione di numerose procedure. Inoltre, in alcune procedure non è possibile utilizzare variabili stringa e altre richiedono valori interi consecutivi per i livelli del fattore.

Figura 8-15
Finestra di dialogo Ricodifica automatica



- Le nuove variabili create dalla procedura Ricodifica automatica mantengono le etichette delle variabili e dei valori definite per la vecchia variabile. Se per alcuni valori non è stata definita un'etichetta dei valori, come etichetta del valore ricodificato verrà utilizzato il valore originale. I valori vecchi e nuovi e le etichette dei valori verranno visualizzati in una tabella.
- I valori stringa verranno ricodificati in ordine alfabetico; le lettere maiuscole precedono le minuscole.
- I valori mancanti verranno ricodificati come maggiori dei valori non mancanti; il loro ordine verrà mantenuto. Se, ad esempio, la variabile originale ha 10 valori non mancanti, il valore mancante minimo verrà ricodificato a 11 e il valore 11 verrà utilizzato come valore mancante per la nuova variabile.

Usa lo stesso schema di ricodifica per tutte le variabili. Questa opzione permette di applicare lo stesso schema di ricodifica a tutte le variabili selezionate e di utilizzare quindi uno schema di codifica uniforme per tutte le nuove variabili.

Se si seleziona questa opzione, si applicano le seguenti regole e limitazioni:

- Tutte le variabili devono essere dello stesso tipo (numeriche o stringa).

- Tutti i valori osservati di tutte le variabili selezionate vengono utilizzate per creare un ordinamento dei valori utilizzabile per la ricodifica nei numeri interi sequenziali.
- I valori mancanti definiti dall'utente delle nuove variabili si basano sulla prima variabile dell'elenco dei valori mancanti definiti dall'utente. Tutti gli altri valori derivanti dalle variabili originali, ad eccezione di quelli mancanti definiti dal sistema, sono considerati validi.

Considera i valori delle stringhe vuote come mancanti definiti dall'utente. Questa opzione può essere usata per i valori delle variabili stringa, vuote o nulle considerate **mancanti di sistema**. Questa opzione ricodifica le stringhe vuote in valori stringa **definiti dall'utente** maggiori del valore mancante massimo.

Modelli

È possibile salvare lo schema di ricodifica in un file di modello e applicarlo a tutte le altre variabili o agli altri file di dati.

Esempio: si devono ricodificare molti codici prodotto alfanumerici in numeri interi ogni mese ed aggiungere, in alcuni mesi, nuovi codici prodotto che modificano lo schema di ricodifica automatica originale. Se si salva lo schema originale in un modello e lo si applica ai nuovi dati che contengono il nuovo insieme di codici, tutti i nuovi codici contenuti nei dati vengono automaticamente ricodificati in valori maggiori dell'ultimo valore del modello, senza che venga modificato lo schema di ricodifica automatica dei codici prodotto originali.

Salva modello con nome. Salva lo schema di ricodifica automatico per le variabili selezionate in un file di modello esterno.

- Questo modello contiene informazioni sulle associazioni tra i valori mancanti originali ed i valori ricodificati.
- Nel modello vengono salvate solo le informazioni relative ai valori non mancanti. Le informazioni relative ai valori non mancanti specificati dall'utente non vengono salvate.

- Se sono state selezionate più variabili per la ricodifica ma non è stato selezionato lo stesso schema di ricodifica automatica per tutte le variabili o non è prevista l'applicazione di un modello esistente allo schema di ricodifica automatica, il modello sarà basato sulla prima variabile dell'elenco.
- Se si selezionano più variabili per la ricodifica e si sceglie l'opzione Usa lo stesso schema di ricodifica per tutte le variabili e/o si seleziona Applica modello, il modello contiene lo schema di ricodifica automatica combinato per tutte le variabili

Applica modello da. Applica un modello di ricodifica automatica già salvato alle variabili selezionate per la ricodifica ed aggiunge tutti i valori aggiuntivi trovati nelle variabili alla fine dello schema, mantenendo la relazione tra i valori originali e ricodificati contenuti nello schema salvato.

- Tutte le variabili selezionate per la ricodifica devono essere dello stesso tipo (numeriche o stringa) ed il tipo deve corrispondere a quello definito nel modello.
- I modelli non contengono informazioni sui valori mancanti definiti dall'utente. I valori mancanti definiti dall'utente delle variabili target sono basati sulla prima variabile dell'elenco delle variabili originali con valori mancanti definiti dall'utente. Tutti gli altri valori derivanti dalle variabili originali, ad eccezione di quelli mancanti definiti dal sistema, sono considerati validi.
- Le associazioni dei valori del modello vengono applicate per prime. Tutti i valori restanti vengono ricodificati in valori superiori all'ultimo valore del modello, mentre i valori mancanti definiti dall'utente (basati sulla prima variabile dell'elenco con valori mancanti definiti dall'utente) vengono ricodificati in valori superiori all'ultimo valore valido.
- Se si selezionano più variabili per la ricodifica automatica, viene applicato prima il modello, seguito da una ricodifica automatica combinata comune per tutti i valori aggiuntivi trovati nelle variabili selezionate. Ciò genera uno schema di ricodifica automatica singolo di tipo comune per tutte le variabili selezionate.

Per ricodificare stringhe o valori numerici in interi consecutivi

▶ Dai menu, scegliere:

Trasforma
  Ricodifica automatica...

▶ Selezionare una o più variabili da ricodificare.

- Per ciascuna variabile selezionata, specificare un nuovo nome di variabile e fare clic su Nuovo nome.

Impostazione guidata data e ora

L'Impostazione guidata data/ora semplifica molte operazioni normalmente associate alle variabili data/ora.

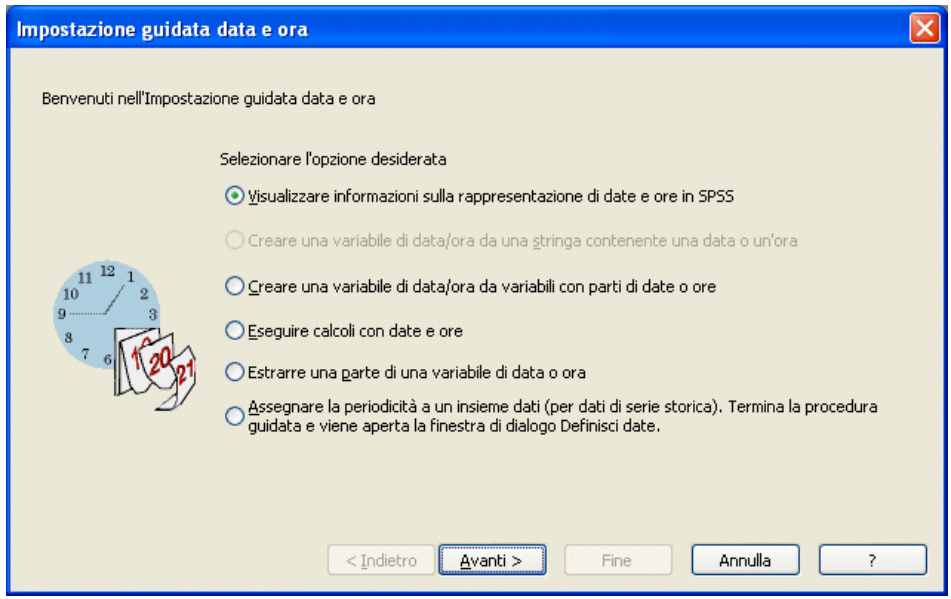
Per utilizzare l'Impostazione guidata data e ora

- Dai menu, scegliere:

Trasforma
  Data/Ora...

- Selezionare l'operazione da eseguire e seguire le istruzioni per definire l'operazione.

Figura 8-16
Schermata introduttiva dell'Impostazione guidata data e ora


Impostazione guidata data e ora
Benvenuti nell'Impostazione guidata data e ora
Selezionare l'opzione desiderata
Visualizzare informazioni sulla rappresentazione di date e ore in SPSS
Creare una variabile di data/ora da una stringa contenente una data o un'ora
Creare una variabile di data/ora da variabili con parti di date o ore
Eseguire calcoli con date e ore
Estrarre una parte di una variabile di data o ora
Assegnare la periodicità a un insieme dati (per dati di serie storica). Termina la procedura guidata e viene aperta la finestra di dialogo Definisci date.
< Indietro
Avanti >
Fine
Annulla
?

- **Informazioni su come vengono rappresentate la data e l'ora in SPSS.** Questa opzione visualizza una schermata che fornisce una breve panoramica sulle variabili data/ora di SPSS. Se si seleziona il pulsante ?, viene visualizzato anche un collegamento con informazioni più dettagliate.
- **Creazione di una variabile di data/ora da una variabile stringa che contiene una data o un orario.** Utilizzare questa opzione per creare una variabile data/ora da una variabile stringa. Ad esempio se la variabile stringa utilizza rappresenta le date nel formato mm/gg/aaaa e si desidera utilizzarla per creare una variabile data/ora.
- **Crea una variabile data/ora dalle variabili che contengono alcuni elementi di date o ore.** Questa opzione permette di creare una variabile data/ora da un insieme di variabili esistenti. Esempio: si supponga di disporre di una variabile che rappresenta il mese (sotto forma di intero), di una seconda variabile che rappresenta il giorno del mese e di una terza variabile che rappresenta l'anno. Le tre variabili possono essere combinate in un'unica variabile data/ora.
- **Eseguire calcoli con date e ore.** Utilizzare questa opzione per aggiungere o sottrarre valori dalle variabili data/ora. Ad esempio è possibile calcolare la durata di un processo sottraendo la variabile che rappresenta l'ora di inizio del processo dalla variabile che rappresenta l'ora di fine.
- **Estrai una parte della variabile data/ora.** Questa opzione permette di estrarre una parte di una variabile di data o ora; ad esempio il giorno del mese da una variabile di data/ora con formato mm/gg/aaaa.
- **Assegna periodicità a file di dati.** Questa opzione consente di visualizzare la finestra di dialogo Definisci date, usata per creare le variabili data/ora costituite da un insieme di date in sequenza. Questa opzione viene generalmente usata per associare date ai dati con serie temporali.

Nota: le attività sono disattivate quando il file di dati non contiene i tipi di variabili richiesti per effettuare l'attività. Ad esempio, se il file di dati contiene variabili non stringa, l'attività di creazione di una variabile data/ora da una stringa non è applicabile ed è disattivata.

Date e ore in SPSS

Le variabili che rappresentano le date e le ore in SPSS sono variabili di tipo numerico, che visualizzano formati che corrispondono ai formati di data/ora specifici. Queste variabili vengono generalmente chiamate variabili data/ora. SPSS distingue tra le variabili data/ora che rappresentano le date e quelle che rappresentano una durata

indipendente da date; ad esempio 20 ore, 10 minuti e 15 secondi. Queste ultime sono esempi di variabili di durata e si distinguono dalle prime che sono variabili data/ora. Per un elenco completo dei formati di visualizzazione, vedere "Data e ora" nella sezione "Formati universali" di *SPSS Command Syntax Reference*.

Variabili data e data/ora. Le variabili data hanno un formato che rappresenta una data; ad esempio mm/gg/aaaa. Le variabili data/ora hanno un formato che rappresenta una data ed un'ora; ad gg-mm-aaaa e hh:mm:ss. Internamente, le variabili data e data/ora vengono salvate in numero di secondi calcolati a partire dal 14 ottobre 1582. Le variabili di data e data/ora sono talvolta chiamate variabile del formato data.

- È possibile specificare l'anno sia con due che con quattro cifre. Per impostazione predefinita, gli anni a due cifre comprendono un intervallo che inizia 69 anni prima e termina 30 anni dopo la data corrente. Questo intervallo varia a seconda delle impostazioni selezionate per Opzioni e può essere configurato selezionando Opzioni e facendo clic su Dati nel menu Modifica.
- I trattini, i punti, le virgole, le barre o gli spazi possono essere usati come delimitatori nei formati giorno-mese-anno.
- I mesi possono essere rappresentati in cifre, numeri romani, abbreviazioni di tre lettere o espressi con il nome completo. Le abbreviazioni di tre lettere o il nome completo devono essere scritti in inglese. I nomi di mesi in altre lingue non vengono riconosciuti.

Variabili durata. Le variabili durata hanno un formato che rappresenta un intervallo di tempo; ad esempio hh:mm. Queste variabili vengono salvate internamente come secondi senza riferimento a una data specifica.

- Nelle specifiche di tempo (applicabili a variabili data/ora e di durata), è possibile usare il punto e virgola per delimitare le ore, i minuti e i secondi. Le ore e i minuti devono sempre essere specificati; i secondi sono opzionali. È sempre necessario inserire un punto per separare i secondi dalle frazioni di secondo. Le ore possono avere qualsiasi grandezza, ma il valore massimo è 59 per i minuti e 59,999.... per i secondi.

Data e ora correnti. La variabile di sistema *$TIME* contiene la data e l'ora correnti. La variabile rappresenta il numero di secondi calcolati dal14 ottobre 1582 alla data e all'ora in cui è stato eseguito il comando di trasformazione.

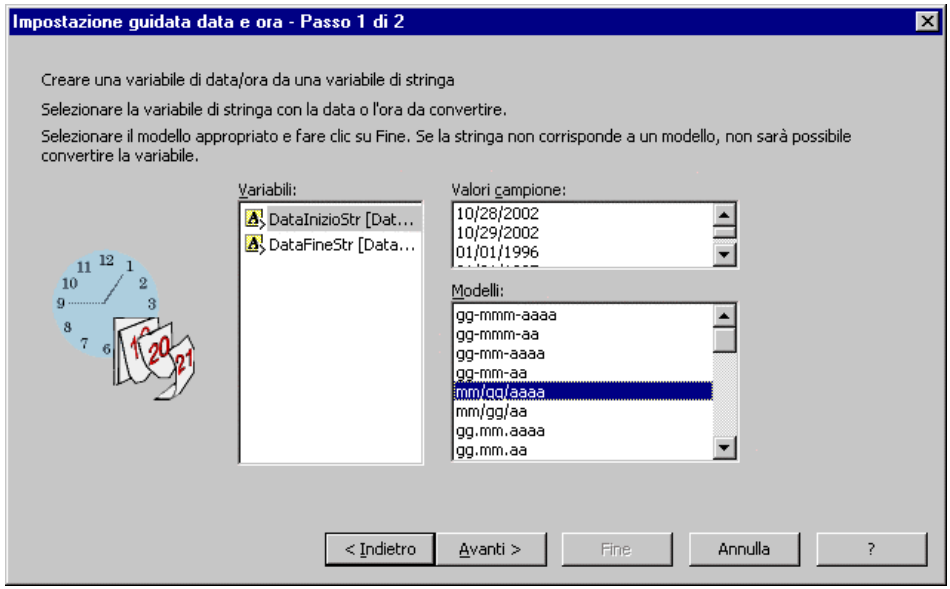
Creazione di una variabile data/ora da una stringa

Per creare una variabile data/ora da una variabile stringa:

- Selezionare Crea una variabile data/ora da una stringa contenente una data o un'ora nella pagina iniziale dell'Impostazione guidata data/ora.

Selezione della variabile stringa da convertire in una variabile data/ora

Figura 8-17
Creazione di una variabile data/ora dalla variabile stringa - Passaggio 1

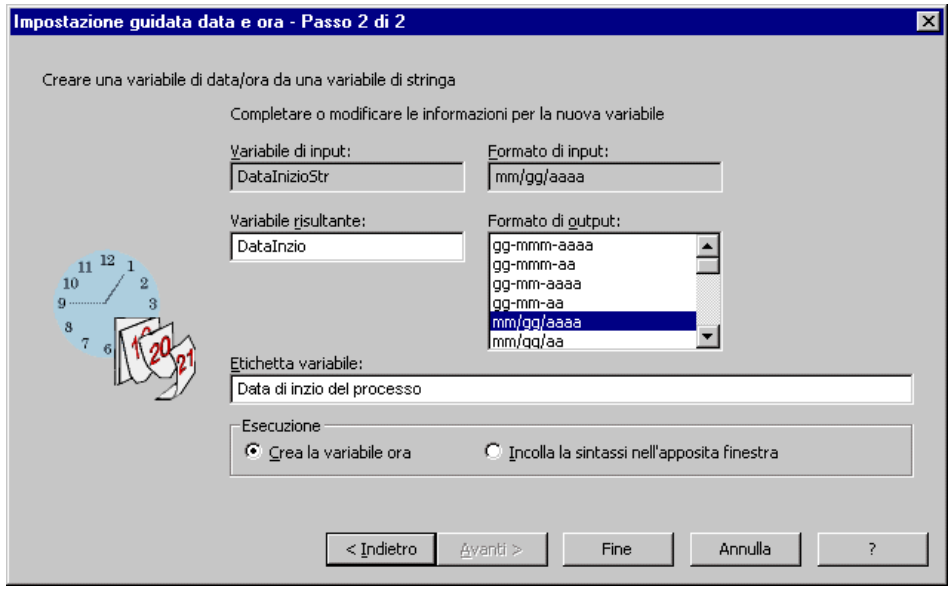


- Selezionare la variabile stringa da convertire nell'elenco Variabili. È utile ricordare che l'elenco comprende solo variabili stringa.

- Selezionare un retino dall'elenco Retino, verificando che corrisponda alla modalità di rappresentazione delle date nella variabile stringa. L'elenco Valori campione visualizza i valori effettivi della variabile selezionata nel file di dati. I valori della variabile stringa che non corrispondono al retino selezionato generano un valore mancante di sistema per la nuova variabile.

Selezione del risultato della conversione della variabile stringa in una variabile data/ora

Figura 8-18
Creazione di una variabile data/ora dalla variabile stringa - Passaggio 2



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

  Se lo si desidera, è possibile:

  - Selezionare un formato data/ora per la nuova variabile nell'elenco Formato dell'output.
  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

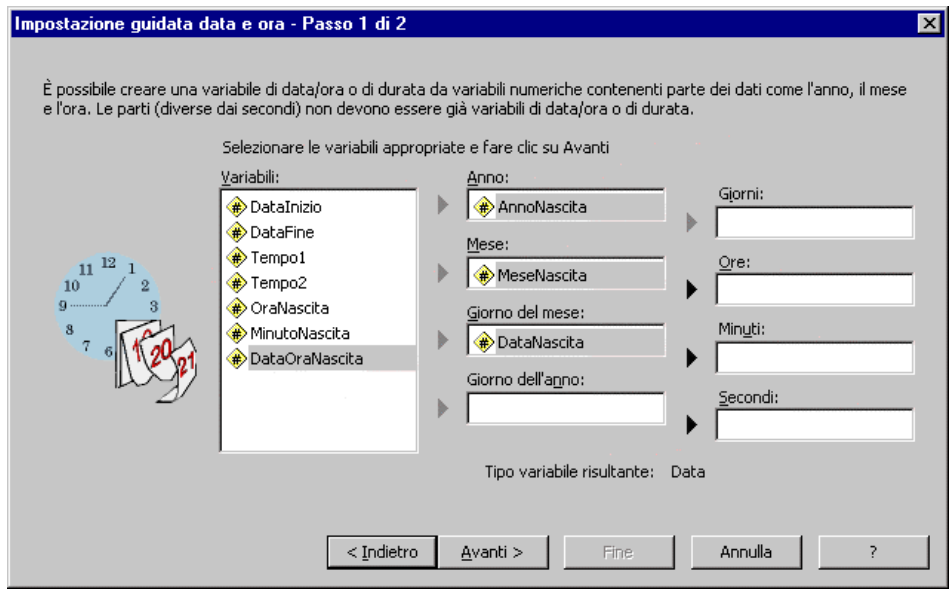
Creazione di una variabile data/ora da un insieme di variabili

Per unire un insieme di variabili esistenti un un'unica variabile data/ora:

- Selezionare Crea una variabile data/ora dalle variabili che contengono parti di date o ore nella pagina iniziale dell'Impostazione guidata data/ora.

Selezione delle variabili da unire alla variabile data/ora

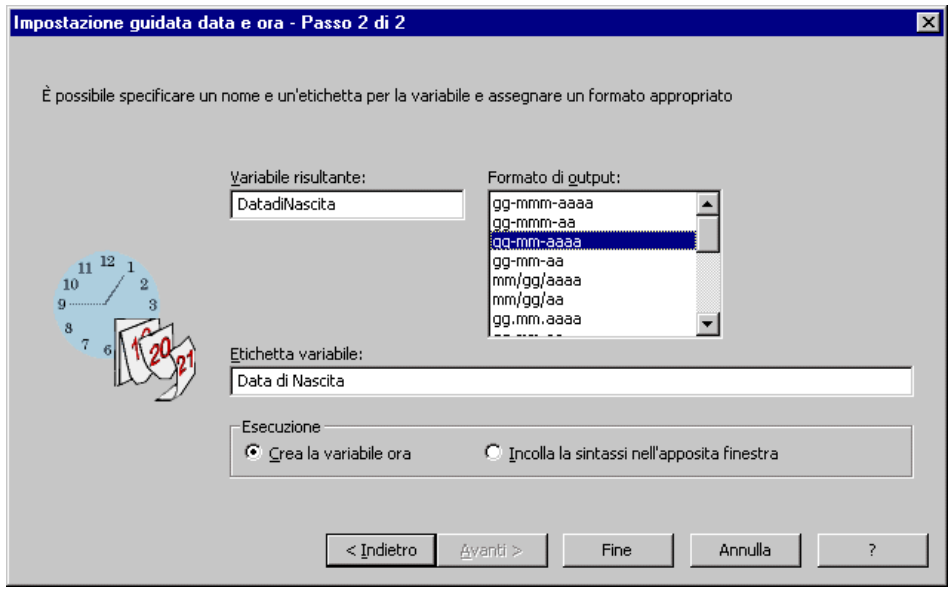
Figura 8-19
Creazione di una variabile data/ora dall'insieme di variabili - Passaggio 1



- Selezionare le variabili che rappresentano le diverse parti della data/dell'ora.
  - Ricordare che non è possibile usare tutte le combinazioni disponibili. Ad esempio non è possibile creare una variabile data/ora da Anno e Giorno del mese perché quando si seleziona Anno è necessario specificare un data completa.
  - Non è possibile usare una variabile data/ora esistente come parte della variabile data/ora finale che si sta creando. Le variabili che costituiscono le parti della nuova variabile data/ora devono essere numeri interi. L'unica eccezione è rappresentata dalla possibilità di usare una variabile data/ora esistente per la parte Secondi della nuova variabile. Poiché è possibile usare le frazioni di secondi, la variabile usata per i Secondi non deve necessariamente essere un numero intero.
  - I valori per qualsiasi parte della nuova variabile che non rientrano nel risultato consentito generano una variabile mancante di sistema per la nuova variabile. Ciò può accadere, ad esempio, se si usa inavvertitamente una variabile che rappresenta un giorno del mese. Poiché l'intervallo valido per i mesi in SPSS è 1–13, a tutti i casi i cui valori dei giorni del mese rientrano nell'intervallo 14–31 verrà assegnato il valore mancante di sistema per la nuova variabile.

Selezione di una variabile data/ora creata mediante l'unione di variabili

Figura 8-20
Creazione di una variabile data/ora dall'insieme di variabili - Passaggio 2



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

- Selezionare un formato di data/ora dall'elenco Formato dell'output.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

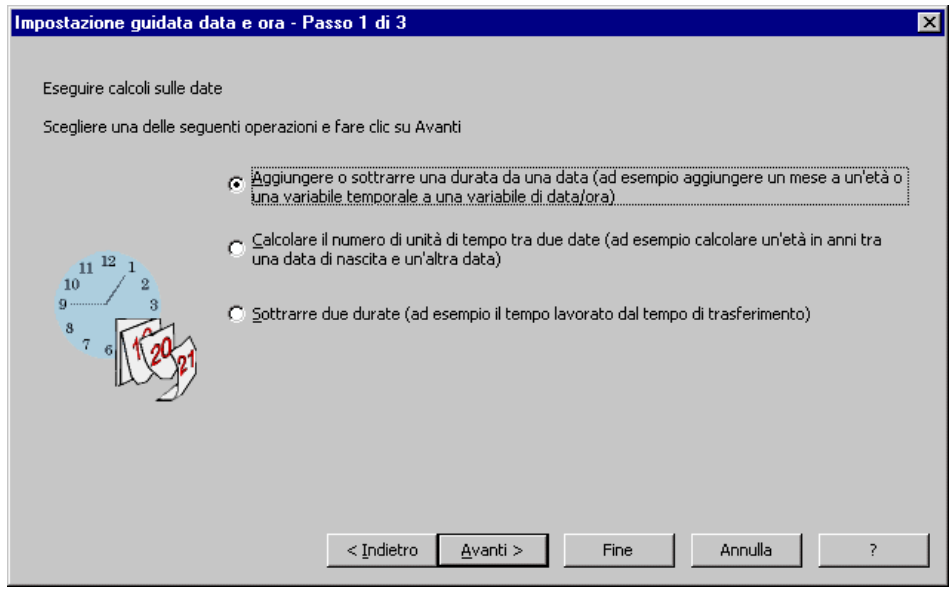
Aggiunta o sottrazione di valori dalle variabili data/ora

Per aggiungere o sottrare valori dalle variabili data/ora:

- Selezionare Eseguire calcoli con date e ore nella pagina iniziale dell'Impostazione guidata data/ora.

Selezione del tipo di calcolo da eseguire con le variabili data/ora

Figura 8-21
Aggiunta o sottrazione di valori alle/dalle variabili data/ora - Passaggio 1


Impostazione guidata data e ora - Passo 1 di 3
Eseguire calcoli sulle date
Scegliere una delle seguenti operazioni e fare clic su Avanti
Aggiungere o sottrarre una durata da una data (ad esempio aggiungere un mese a un'età o una variabile temporale a una variabile di data/ora)
Calcolare il numero di unità di tempo tra due date (ad esempio calcolare un'età in anni tra una data di nascita e un'altra data)
Sottrarre due durate (ad esempio il tempo lavorato dal tempo di trasferimento)
< Indietro
Avanti >
Fine
Annulla
?

- **Aggiunti o sottrai una durata alla/dalla data.** Utilizzare questa opzione per aggiungere o sottrarre a una/da una variabile formato data. È possibile aggiungere o sottrare durate corrispondenti a valori fissi come 10 giorni oppure valori da una variabile numerica; ad esempio una variabile che rappresenta gli anni.
- **Calcola il numero di unità di tempo tra due date.** Usare questa opzione per ottenere la differenza tra due date nell'unità selezionata. Ad esempio è possibile ottenere il numero di anni o il numero di giorni che separano due date.
- **Sottrai due durate.** Usare questa opzione per ottenere la differenza tra due variabili che hanno formati di durata; ad esempio hh:mm o hh:mm:ss.

Nota: le attività sono disattivate quando il file di dati non contiene i tipi di variabili richiesti per effettuare l'attività. Ad esempio se il file di dati non contiene due variabili con formati di durata, l'operazione di sottrazione delle due durate non è applicabile e viene disattivata.

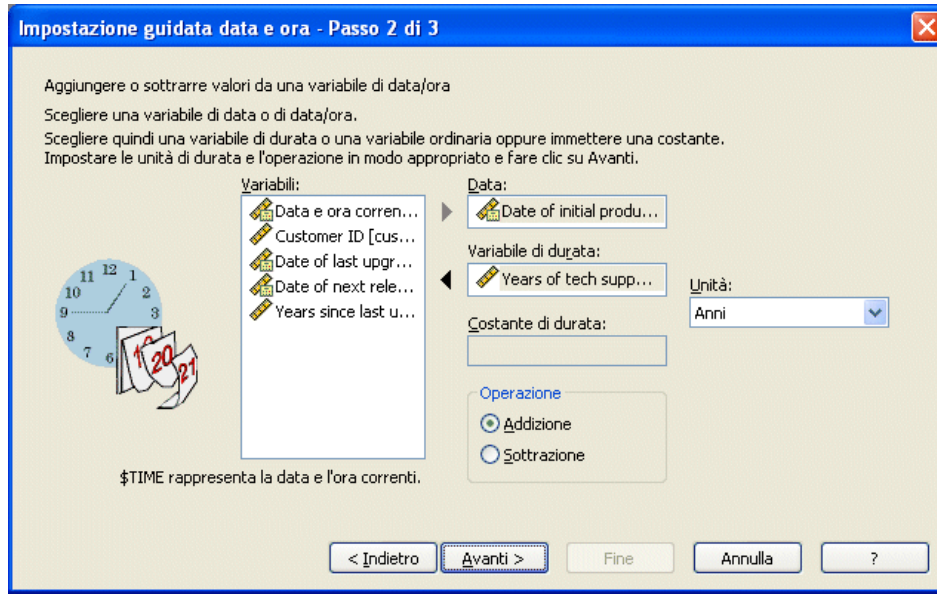
Aggiunta o sottrazione di una durata a/da una data

Per aggiungere o sottrarre una durata da una variabile formato data:

- Selezionare Aggiungi o sottrai una durata a/da una data nella schermata Esegui calcoli sulle date dell'Impostazione guidata data/ora.

Selezione della variabile data/ora e della durata da aggiungere o sottrarre

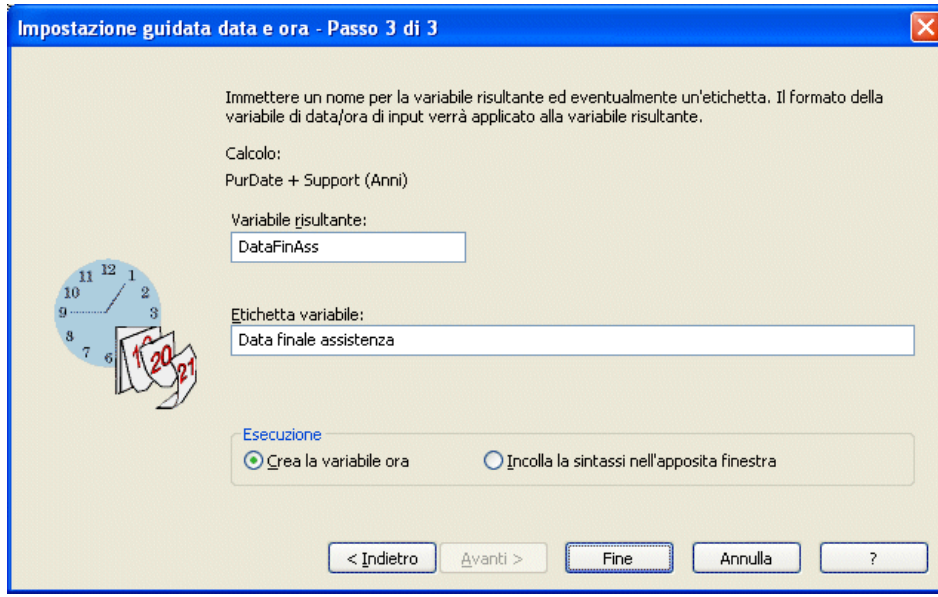
Figura 8-22
Aggiunta o sottrazione della durata - Passaggio 2


Impostazione guidata data e ora - Passo 2 di 3
Aggiungere o sottrarre valori da una variabile di data/ora
Scegliere una variabile di data o di data/ora.
Scegliere quindi una variabile di durata o una variabile ordinaria oppure immettere una costante.
Impostare le unità di durata e l'operazione in modo appropriato e fare clic su Avanti.
Variabili:
Data e ora corren...
Customer ID [cus...
Date of last upgr...
Date of next rele...
Years since last u...
Data:
Date of initial produ...
Variabile di durata:
Years of tech supp...
Unità:
Anni
Costante di durata:
Operazione
Addizione
Sottrazione
$TIME rappresenta la data e l'ora correnti.
< Indietro
Avanti >
Fine
Annulla
?

- Selezionare una variabile data (o ora).
- Selezionare una variabile durata oppure immettere un valore in Costante durata. Le variabili usate per le durate non possono essere variabili data o data/ora. Tuttavia, possono essere variabili di durata o semplici variabili numeriche.
- Selezionare l'unità della durata dall'elenco a discesa. Selezionare Durata se si utilizza una variabile di durata come hh:mm o hh:mm:ss.

Selezione del risultato dell'aggiunta o sottrazione di una durata a/da una variabile data/ora

Figura 8-23
Aggiunta o sottrazione della durata - Passaggio 3


Impostazione guidata data e ora - Passo 3 di 3
Immettere un nome per la variabile risultante ed eventualmente un'etichetta. Il formato della variabile di data/ora di input verrà applicato alla variabile risultante.
Calcolo:
PurDate + Support (Anni)
Variabile risultante:
DataFinAss
Etichetta variabile:
Data finale assistenza
Esecuzione
Crea la variabile ora
Incolla la sintassi nell'apposita finestra
< Indietro
Avanti >
Fine
Annulla
?

- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

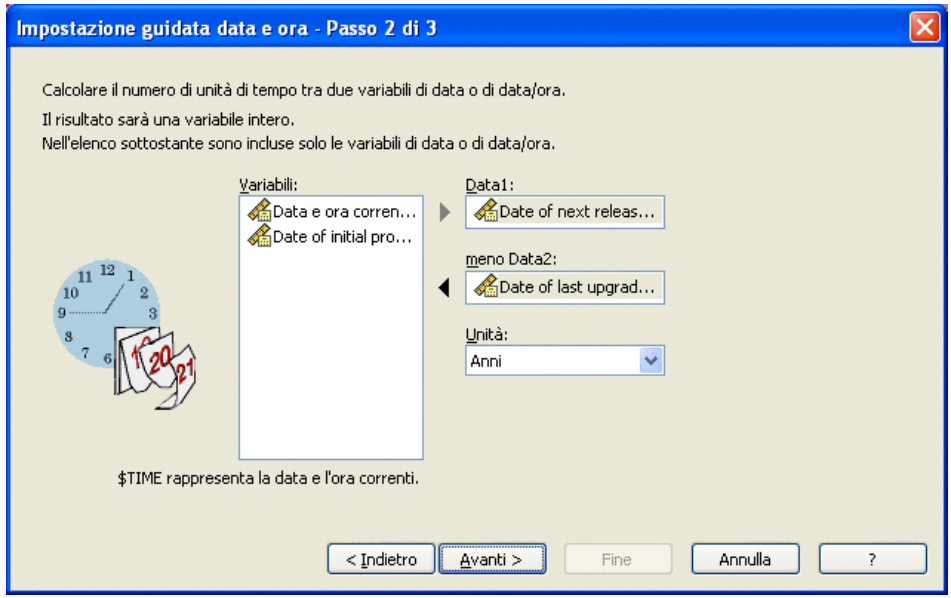
Sottrazione di variabili in formato data

Per sottrarre due variabili in formato data:

- Selezionare Calcola numero di unità di tempo tra due date nella schermata Esegui calcoli sulle date dell'Impostazione guidata data/ora.

Selezione delle variabli in formato data da sottrarre

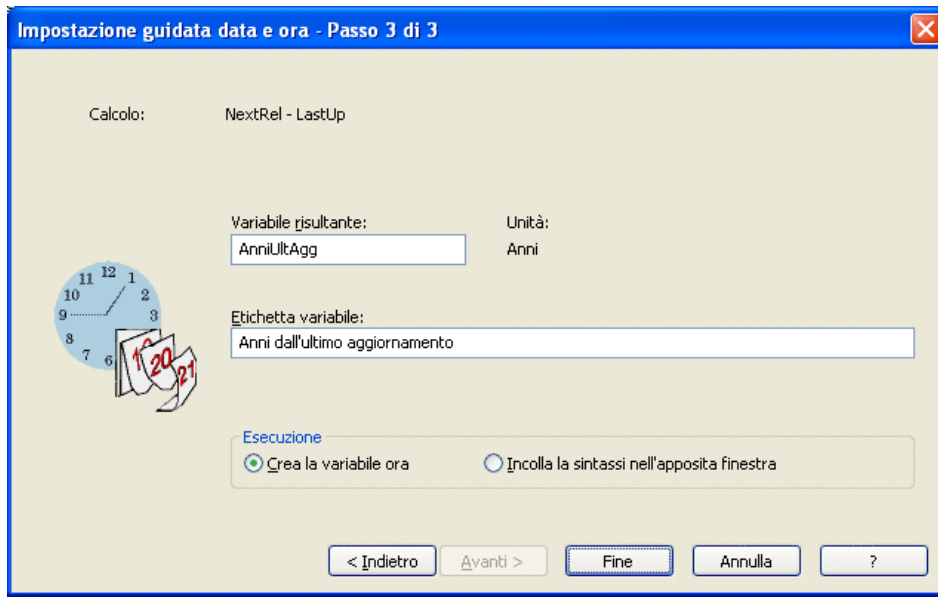
Figura 8-24
Sottrazione di date - Passaggio 2


Impostazione guidata data e ora - Passo 2 di 3
Calcolare il numero di unità di tempo tra due variabili di data o di data/ora.
Il risultato sarà una variabile intero.
Nell'elenco sottostante sono incluse solo le variabili di data o di data/ora.
Variabili:
Data e ora corren...
Date of initial pro...
Data1:
Date of next releas...
meno Data2:
Date of last upgrad...
Unità:
Anni
$TIME rappresenta la data e l'ora correnti.
< Indietro
Avanti >
Fine
Annulla
?

- Selezionare le variabili da sottrarre.
- Selezionare l'unità per il risultato dall'elenco a discesa.

Selezione del risultato per la sottrazione di due variabili in formato data

Figura 8-25
Sottrazione di date - Passaggio 3



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

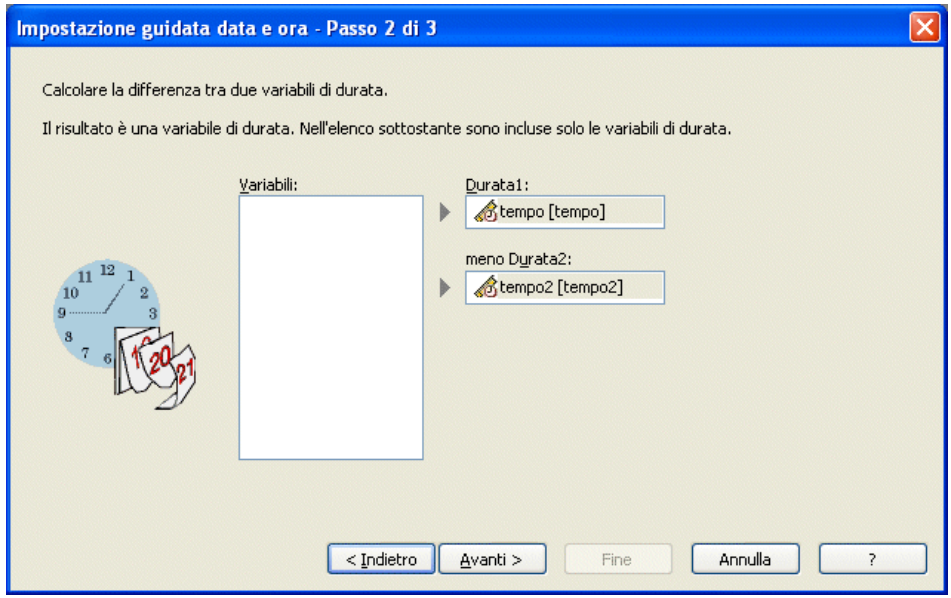
Sottrazione di variabili di durata

Per sottrarre due variabili di durata:

- Selezionare Sottrai due durate nella schermata Esegui calcoli sulle date dell'Impostazione guidata data/ora.

Selezione delle variabili di durata da sottrarre

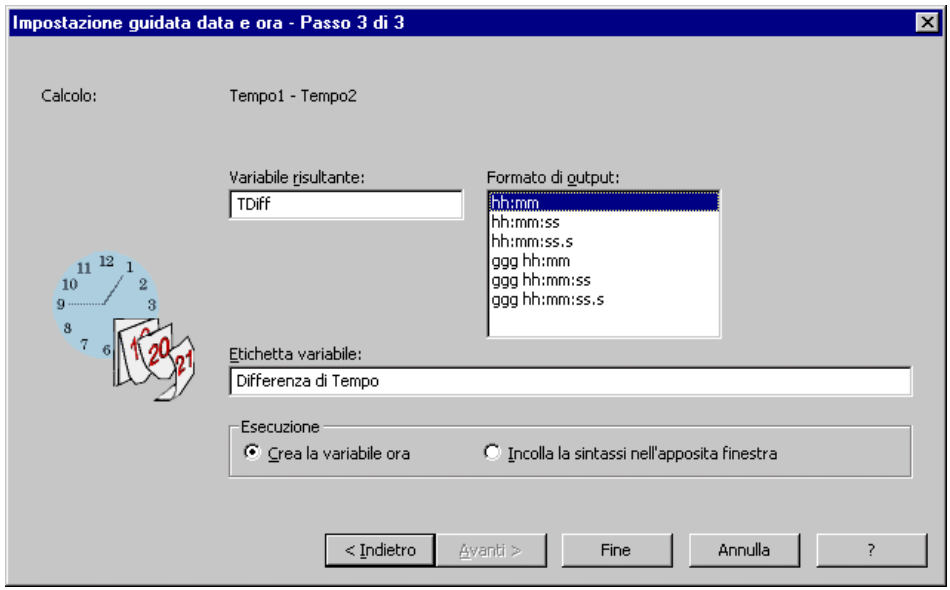
Figura 8-26
Sottrazione di durate - Passaggio 2


Impostazione guidata data e ora - Passo 2 di 3
Calcolare la differenza tra due variabili di durata.
Il risultato è una variabile di durata. Nell'elenco sottostante sono incluse solo le variabili di durata.
Variabili:
Durata1:
tempo [tempo]
meno Durata2:
tempo2 [tempo2]
< Indietro
Avanti >
Fine
Annulla
?

- Selezionare le variabili da sottrarre.

Selezione del risultato della sottrazione di due variabili di durata

Figura 8-27
Sottrazione di durate - Passaggio 3



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.

- Selezionare un formato di durata dall'elenco Formato dell'output.

  Se lo si desidera, è possibile:

  - Assegnare un'etichetta variabile descrittiva alla nuova variabile.

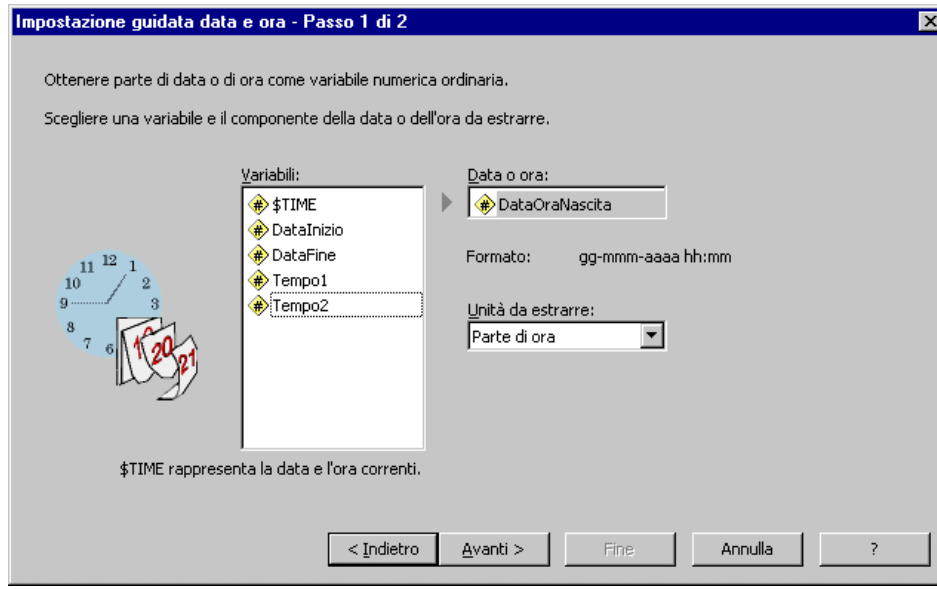
Estrazione di parte della variabile data/ora

Per estrarre un componente, come l'anno, da una variabile data/ora:

- Selezionare Estrai una parte della variabile data o ora nella schermata iniziale dell'Impostazione guidata data/ora.

Selezione del componente da estrarre dalla variabile data/ora

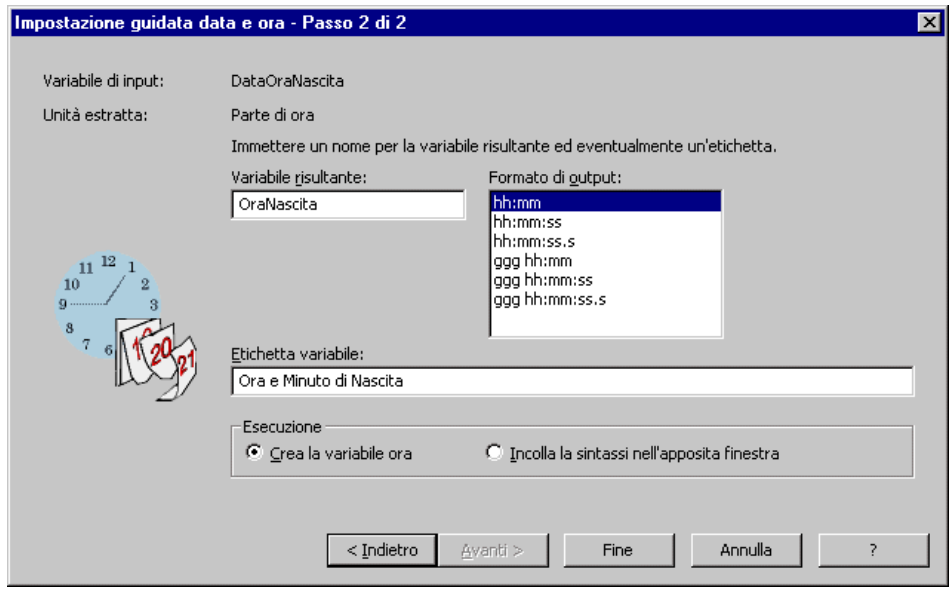
Figura 8-28
Acquisizione della parte della variabile data/ora - Passaggio 1



- Selezionare la variabile che contiene la parte di data o ora da estrarre.
- Selezionare la parte della variabile da estrarre dall'elenco a discesa. È possibile estrarre informazioni dalle date anche se non fanno esplicitamente parte della data visualizzata; come qualsiasi giorno della settimana.

Selezione del risultato per l'estrazione del componente dalla variabile data/ora

Figura 8-29
Acquisizione della parte della variabile data/ora - Passaggio 2



- Specificare un nome per la variabile del risultato. Il nome non può corrispondere a quello di una variabile esistente.
- Se si sta estraendo la parte data o ora da una variabile data/ora, è necessario selezionare un formato dall'elenco Formato dell'output. Questo elenco è disattivato per i casi in cui non è richiesto.

Se lo si desidera, è possibile:

- Assegnare un'etichetta variabile descrittiva alla nuova variabile.

Trasformazioni di dati di serie storiche

Sono disponibili numerose trasformazioni di dati che risultano utili per l'analisi delle serie storiche:

- Generazione di variabili della data per definire la periodicità e distinguere tra periodi storici, di convalida e di previsione.

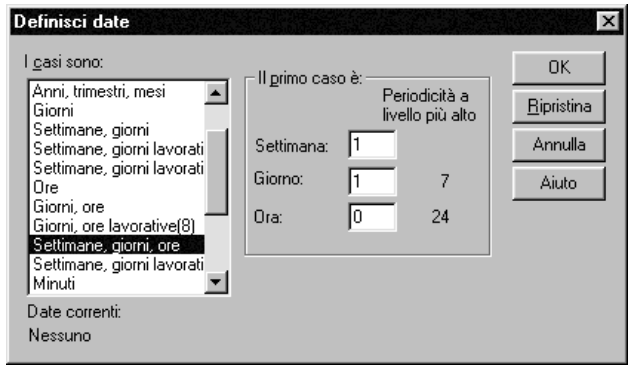
- Creazione di nuove variabili di serie storica come funzioni di variabili di serie storica esistenti.
- Sostituzione di valori mancanti di sistema e definiti dall'utente con stime basate su diversi metodi.

Per ottenere una **serie storica** è necessario misurare una variabile (o un insieme di variabili) regolarmente per un periodo di tempo. Le trasformazioni di dati di serie storiche si fondano su una struttura del file di dati in cui ciascun caso (riga) rappresenta una serie di osservazioni eseguite in momenti diversi e con un periodo di tempo uniforme tra i casi.

Definisci date

La finestra di dialogo Definisci date permette di generare variabili data utilizzabili per definire la periodicità di **serie storiche** e per etichettare l'output di un'analisi di serie storiche.

Figura 8-30
Finestra di dialogo Definisci date



I casi sono. Consente di definire l'intervallo di tempo utilizzato per generare le date.

- Senza data Consente di rimuovere le variabili della data precedentemente definite. Tutte le variabili con i seguenti nomi vengono eliminate: *year_*, *quarter_*, *month_*, *week_*, *day_*, *hour_*, *minute_*, *second_*, and *date_*.
- Personalizzato indica la presenza di variabili della data personalizzate create con la sintassi (ad esempio, una settimana lavorativa di quattro giorni). Questo riflette semplicemente lo stato del file dati di lavoro attivo. Selezionandolo dall'elenco non si ottiene alcun risultato.

Il primo caso è. Consente di definire il valore iniziale che verrà assegnato al primo caso. Ai valori successivi verranno assegnati valori sequenziali in base all'intervallo di tempo.

Periodicità al livello più alto. Indica la variazione ciclica ripetitiva, ad esempio il numero di mesi dell'anno o il numero di giorni della settimana. Il valore visualizzato indica il massimo valore che è possibile inserire.

Per ciascun componente utilizzato per definire la data verrà creata una nuova variabile numerica. I nomi delle nuove variabili terminano con un carattere di sottolineatura. Per i componenti verrà inoltre creata una variabile stringa descrittiva, *date_*. Se, ad esempio, si seleziona Settimane, giorni, ore, verranno create quattro nuove variabili: *week_*, *day_*, *hour_* e *date_*.

Se i nomi delle variabili sono già stati definiti, verranno sostituiti quando si creano nuove variabili con gli stessi nomi delle variabili della data esistenti.

Per definire date per dati di serie storiche

- Dai menu, scegliere:
  Dati
  Definisci date...
- Selezionare un intervallo di tempo per l'elenco I casi sono.
- Specificare uno o più valori che definiscano la data iniziale per Il primo caso è, che determinerà la data assegnata al primo caso.

Variabili della data e variabili del formato della data

Le variabili della data create con l'opzione Definisci date non devono essere confuse con le variabili del formato della data, definite con l'opzione Visualizzazione variabili dell'Editor dei dati. Le variabili della data vengono utilizzate per stabilire la periodicità dei dati di serie storiche. Le variabili del formato della data rappresentano date e/o orari espressi in diversi formati. Le variabili della data sono interi semplici che rappresentano, ad esempio, il numero di giorni, settimane, ore a partire da un punto iniziale definito dall'utente. La maggior parte delle variabili del formato della data vengono memorizzate internamente come il numero di secondi trascorsi dal 14 ottobre 1582.

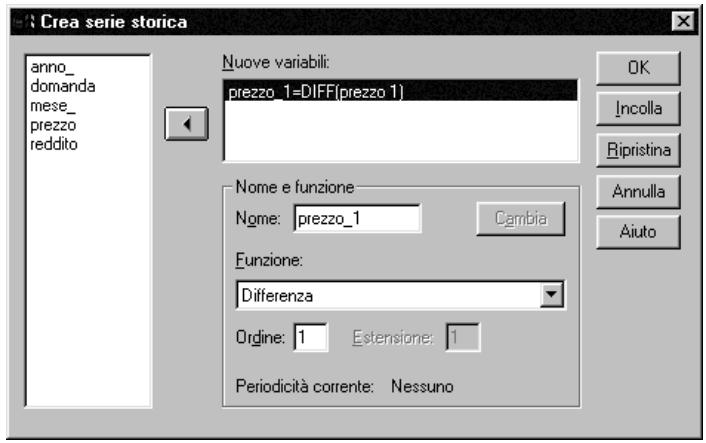
Crea serie storica

La finestra di dialogo Crea serie storica permette di creare nuove variabili basate su funzioni di variabili numeriche di **serie storiche** variabili. Questi valori trasformati risultano utili in numerose procedure di analisi delle serie storiche.

I nuovi nomi di variabile predefiniti sono rappresentati dai primi sei caratteri della variabile esistente in base alla quale sono stati creati, seguiti da un carattere di sottolineatura e da un numero sequenziale. Ad esempio, per la variabile *prezzo*, il nuovo nome di variabile sarà *prezzo_1*. Le nuove variabili mantengono le etichette dei valori definite per le variabili esistenti.

Per la creazione di variabili di serie storica sono disponibili diverse funzioni, ad esempio differenze, medie mobili, mediane mobili, ritardo e anticipo.

Figura 8-31
Finestra di dialogo Crea serie storica



Per creare nuove variabili di serie storica

- Dai menu, scegliere:

  Trasforma
  Crea serie storica...

- Selezionare la funzione di serie storica che si desidera utilizzare per trasformare le variabili originali.

- Selezionare le variabili in base alle quali creare nuove variabili di serie storica. È possibile utilizzare solo variabili numeriche.

Se lo si desidera, è possibile:

- Specificare nomi di variabili per annullare i nomi predefiniti delle nuove variabili.
- Modificare la funzione relativa alla variabile selezionata.

Funzioni di trasformazione delle serie storiche

Differenza. Differenza non stagionale tra valori successivi della serie. L'ordine è rappresentato dal numero di valori precedenti utilizzati per calcolare la differenza. Poiché per ciascun ordine di differenza va perduta un'osservazione, i valori mancanti di sistema compaiono all'inizio della serie. Se, ad esempio, l'ordine della differenza è 2, ai primi due casi verrà assegnato il valore mancante di sistema per la nuova variabile.

Differenza stagionale. Differenza tra i valori della serie basata su un intervallo di tempo costante. L'intervallo viene definito in base alla periodicità corrente. Per calcolare le differenze stagionali, è innanzitutto necessario definire variabili della data (menu Dati, Definisci dati) che includano un componente periodico (ad esempio i mesi dell'anno). L'ordine è rappresentato dal numero di periodi stagionali utilizzati per calcolare la differenza. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale alla periodicità moltiplicata per l'ordine. Se, ad esempio, la periodicità corrente è 12 e l'ordine è 2, ai primi 24 casi verrà assegnato il valore mancante di sistema per la nuova variabile.

Media mobile centrata. Media di un intervallo di valori di serie che circondano e includono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie utilizzati per calcolare la media. Se l'intervallo è pari, la media mobile verrà calcolata in base alla media di ciascuna coppia di medie non centrate. Il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie per un intervallo di n equivale a $n/2$ per gli intervalli di valori pari e dispari. Se, ad esempio, l'intervallo è 5, il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie è 2.

Media mobile sui precedenti. Media dell'intervallo di valori di serie che precedono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie precedenti utilizzati per calcolare la media. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale al valore dell'intervallo.

Mediana mobile. Mediane di intervalli di valori di serie che circondano e includono il valore corrente. L'intervallo è rappresentato dal numero di valori di serie utilizzati per calcolare la mediana. Se l'intervallo è pari, la mediana verrà calcolata in base

alla media di ciascuna coppia di mediane non centrate. Il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie per un intervallo di *n* equivale a *n*/2 per gli intervalli di valori pari e dispari. Se, ad esempio, l'intervallo è 5, il numero di casi con il valore mancante di sistema all'inizio e alla fine della serie è 2.

Somma cumulata. La somma cumulata dei valori di serie che precedono e includono il valore corrente.

Ritardo. Valore di un caso precedente, calcolato in base all'ordine di ritardo specificato. L'ordine è rappresentato dal numero di casi che precedono il valore corrente da cui si ottiene tale valore. Il numero di casi con il valore mancante di sistema all'inizio della serie equivale al valore dell'ordine.

Anticipo. Valore di un caso successivo, calcolato in base all'ordine di anticipo specificato. L'ordine è rappresentato dal numero di casi che seguono il caso corrente da cui si ottiene tale valore. Il numero di casi con il valore mancante di sistema alla fine della serie equivale al valore dell'ordine.

Livellamento. Nuovi valori di serie calcolati in base a un algoritmo di livellamento. Il livellatore inizia con la mediana mobile 4, centrata in base alla mediana mobile 2. I valori vengono quindi livellati nuovamente applicando la mediana mobile 5, la mediana mobile 3 e le medie mobili ponderate. I residui vengono calcolati sottraendo le serie livellate dalla serie originale. L'intero processo viene quindi ripetuto sui residui calcolati. Infine, vengono calcolati i residui livellati sottraendo i valori livellati ottenuti nella prima fase del processo. Questo procedimento talvolta viene definito livellamento T4253H.

Sostituisci valori mancanti

Le osservazioni mancanti possono creare problemi durante l'analisi e non è possibile calcolare alcune serie storiche misurate se queste includono valori mancanti. In alcuni casi il valore di un'osservazione specifica non è noto. Inoltre, i dati mancanti possono derivare da:

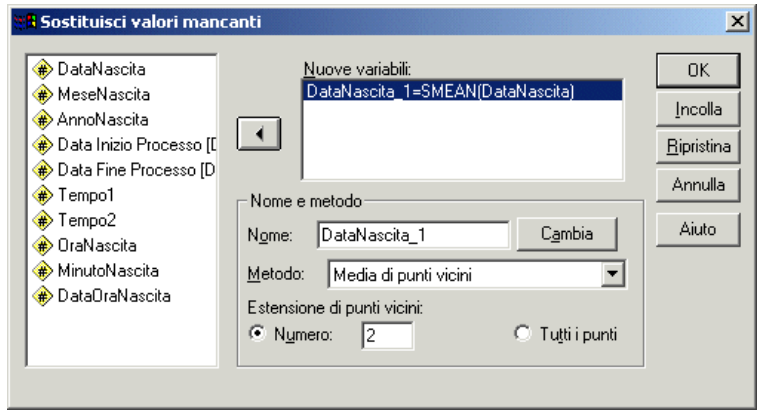
- Ciascun grado di differenziazione riduce la lunghezza della serie di 1.
- Ciascun grado di differenziazione stagionale riduce la lunghezza della serie di una stagione.

- Se si crea una nuova serie contenente previsioni che si estendono oltre la serie esistente (facendo clic su Salva e selezionando opzioni idonee), la serie originale e la serie residua risultante conterranno dati mancanti per le nuove osservazioni.
- Alcune trasformazioni come quella logaritmica generano dati mancanti per alcuni valori della serie originale.

I dati mancanti all'inizio o alla fine della serie non pongono problemi, perché riducono semplicemente la lunghezza utile della serie. Intervalli vuoti al centro di una serie (dati mancanti *incorporati*) può porre problemi molto più seri. La portata del problema dipende dal tipo di procedura di analisi usato.

La finestra di dialogo Sostituisci valori mancanti permette di creare nuove variabili di **serie storiche** utilizzando quelle esistenti nonché sostituire i valori mancanti con le stime calcolate con metodi diversi. I nuovi nomi di variabile predefiniti sono rappresentati dai primi sei caratteri della variabile esistente in base alla quale sono stati creati, seguiti da un carattere di sottolineatura e da un numero sequenziale. Ad esempio, per la variabile *prezzo*, il nuovo nome di variabile sarà *prezzo_1*. Le nuove variabili mantengono le etichette dei valori definite per le variabili esistenti.

Figura 8-32
Finestra di dialogo Sostituisci valori mancanti



Per sostituire i valori mancanti per le variabili di serie storica

▶ Dai menu, scegliere:

Trasforma
  Sostituisci valori mancanti...

- Selezionare il metodo di stima che si desidera utilizzare per la sostituzione dei valori mancanti.
- Selezionare le variabili per cui si desidera sostituire i valori mancanti.

  Se lo si desidera, è possibile:

  - Specificare nomi di variabili per annullare i nomi predefiniti delle nuove variabili.
  - Modificare il metodo di stima per la variabile selezionata.

Metodi di stima per la sostituzione dei valori mancanti

Media della serie. Consente di sostituire i valori mancanti con la media dell'intera serie.

Media di punti vicini. Consente di sostituire i valori mancanti con la media dei valori validi circostanti. L'intervallo di punti vicini è il numero di valori validi situati sopra e sotto il valore mancante utilizzato per calcolare la media.

Mediana di punti vicini. Consente di sostituire i valori mancanti con la mediana dei valori validi circostanti. L'intervallo dei punti vicini è il numero di valori validi situati sopra e sotto il valore mancante utilizzato per calcolare la mediana.

Interpolazione lineare. Consente di sostituire i valori mancanti utilizzando l'interpolazione lineare. Per l'interpolazione verranno utilizzati l'ultimo valore valido prima del valore mancante e il primo valore valido dopo il valore mancante. Se il primo o l'ultimo caso della serie contengono un valore mancante, quest'ultimo non verrà sostituito.

Trend lineare in quel punto. Consente di sostituire i valori mancanti con il trend lineare per quel punto. Sulla serie esistente verrà eseguita una regressione su una variabile indice che assume i valori da 1 a *n*. I valori mancanti verranno sostituiti con i rispettivi valori attesi.

Calcolo del punteggio dei dati con modelli predittivi

Il processo di applicare un modello predittivo ad un insieme di dati viene definito **calcolo del punteggio** dei dati. SPSS, Clementine e AnswerTree offrono procedure che consentono di creare modelli predittivi come i modelli di regressione, clustering, albero e delle reti neurali. Dopo la creazione di un modello, le sue specifiche possono essere salvate come file XML che contiene tutte le informazioni necessarie per

ricostruire il modello. Il prodotto SPSS Server fornisce il sistema per leggere il file modello XML ed applicarlo al file di dati.

Esempio. Un'applicazione di credito viene utilizzata per valutare i rischi connessi alla concessione di un prestito. Il punteggio del credito che si ottiene dal modello di rischio può essere usato per decidere se concedere o meno il prestito.

Il calcolo del punteggio è equivalente ad una trasformazione dei dati. Il modello viene espresso internamente come un insieme di trasformazioni numeriche che possono essere applicate ad un insieme specifico di variabili - le variabili predittore del modello - per ottenere il risultato previsto. In questo senso il processo di calcolare il punteggio con un modello specifico è equivalente ad applicare una funzione, come una radice quadrata, ad un insieme di dati.

Il calcolo del punteggio è disponibile solo con SPSS Server e può essere eseguito interattivamente dagli utenti che utilizzano la modalità di analisi distribuita. Per calcolare il punteggio di file di dati di grandi dimensioni, è consigliabile ricorrere all'utilità Batch di SPSS, una versione eseguibile separata di SPSS fornita con SPSS Server. Per informazioni su come usare l'utilità SPSS Batch Facility, vedere *SPSS Batch Facility User's Guide*, fornito come file PDF sul CD-ROM che contiene SPSS Server.

Il processo di calcolo del punteggio include quanto segue:

- Caricamento di un modello da un file in formato XML (PMML).
- Calcolo del punteggio come nuova variabile, utilizzando la funzione ApplyModel o StrApplyModel disponibile nella finestra di dialogo Calcola variabile.

Per informazioni dettagliate sulla funzione `ApplyModel` o `StrApplyModel`, vedere le espressioni per il calcolo dei punteggi nella sezione Espressioni di trasformazione di *SPSS Command Syntax Reference*.

Caricamento di un modello salvato

La finestra di dialogo Carica modello, che consente di caricare modelli predittivi salvati in formato XML (PMML), è disponibile solo quando si utilizza la modalità di analisi distribuita. ll caricamento dei modelli è il primo passo necessario per poter utilizzare tali modelli per il calcolo del punteggio dei dati.

Figura 8-33
Output di Carica modello

mregression

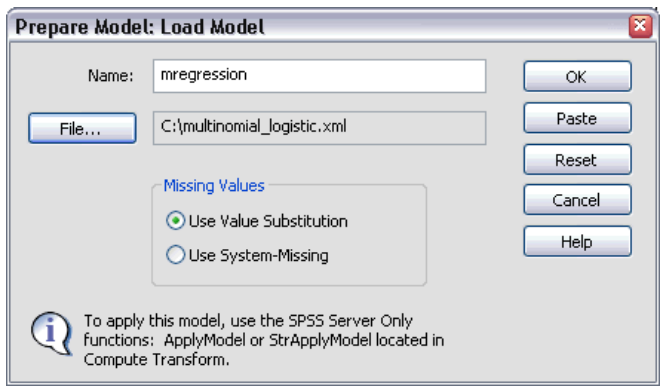
SPSS Variables	Model Variables					
	Name	Label	Type	Width	Role	Measurement
Age_Group	Age_ Group	Age_ Group	Numeric	8	Factor	Nominal
Frequency	Frequency	Frequency	Numeric	7	Covariate	Continuous
Gender	Gender	Gender	String	6	Factor	Nominal
Has_Broadband	Has_ Broadband	Has_ Broadband	Numeric	10	Factor	Nominal
Has_Child	Has_Child	Has_Child	Numeric	10	Factor	Nominal
Income_Group	Income_ Group	Income_ Group	Numeric	10	Factor	Nominal
Log_Amount	Log_ Amount	Log_ Amount	Numeric	8	Covariate	Continuous
Recency	Recency	Recency	Numeric	10	Covariate	Continuous
Response	Response	Response	Numeric	10	Target	Nominal

Per caricare un modello

▶ Dai menu, scegliere:

Trasforma
  Prepara modello
    Carica modello...

Figura 8-34
Finestra di dialogo Carica modello: Prepara modello



▶ Specificare un nome da associare a questo modello. A ogni modello caricato è necessario assegnare un nome univoco.

▶ Fare clic su File e selezionare un file modello. La finestra di dialogo Apri file che viene visualizzata indica i file disponibili in modalità di analisi distribuita. Sono inclusi i file presenti nel sistema in cui è installato SPSS Server e i file memorizzati nel computer locale che risiedono in cartelle o unità condivise.

Nota: quando si calcolano i punteggi dei dati, il modello verrà applicato alle variabili del file di dati attivo che possiedono lo stesso nome delle variabili presenti nel file modello. È possibile associare le variabili del modello originale a diverse variabili del file di dati attivo utilizzando la sintassi dei comandi (vedere il comando `MODEL HANDLE`).

Nome. Nome utilizzato per identificare il modello. Le regole relative ai nomi di modello validi sono le stesse che vengono applicate ai nomi di variabili SPSS (vedere "Nomi delle variabili"Capitolo 5 a pag. 99), con l'aggiunta del carattere $ come primo carattere ammesso. Questo nome potrà essere utilizzato per specificare il modello durante il calcolo del punteggio dei dati con la funzione ApplyModel o StrApplyModel.

File. File XML (PMML) contenente le specifiche del modello.

Valori mancanti

Questo gruppo di opzioni controlla il trattamento dei valori mancanti, incontrati durante il processo di calcolo del punteggio, per le variabili predittive definite nel modello. Un valore mancante nel contesto del calcolo del punteggio fa riferimento a uno dei seguenti elementi:

- Le variabili predittive non contengono alcun valore. Nel caso delle variabili numeriche, questo indica il valore mancante di sistema e nelle variabili di stringa una stringa nulla.
- Nel modello il valore è stato definito come valore utente non valido per il predittore dato. I valori definiti come valori utente non validi nel file di dati attivo, ma non nel modello, non vengono identificati come valori mancanti nel processo di calcolo.
- La variabile predittiva è categoriale e il valore non rappresenta una delle categorie definite nel modello.

Usa valore di sostituzione. Tenta di utilizzare la sostituzione del valore durante il calcolo del punteggio nei casi che presentano valori mancanti. Il metodo per determinare quale valore utilizzare per sostituire un valore mancante dipende dal tipo di modello predittivo.

- **Modelli SPSS.** Per quanto riguarda le variabili indipendenti nei modelli di regressione lineare e discriminanti, se durante la creazione e il salvataggio del modello è stato specificato il valore di sostituzione medio per i valori mancanti, tale valore verrà utilizzato in luogo dei valori mancanti nel calcolo del punteggio, in modo da consentire la continuazione del calcolo stesso. Se il valore medio non è disponibile, verrà restituito il valore mancante di sistema.
- **Modelli AnswerTree & Modelli di comando SPSS TREE.** Nei modelli CHAID e CHAID esaustivo, per la variabile di distinzione mancante viene selezionato il nodo figlio più grande, ovvero il nodo figlio con la popolazione più estesa tra i nodi figlio che utilizzano casi di esempio per l'apprendimento. Nei modelli C&RT e QUEST le variabili di distinzione surrogate (se presenti) vengono utilizzate per prime (le suddivisioni surrogate sono suddivisioni che tentano di corrispondere il più possibile alle suddivisioni originali ricorrendo a predittori alternativi). Se non si specifica alcuna suddivisione surrogata o se tutte le variabili di distinzione surrogate risultano mancanti, verrà utilizzato il nodo figlio più grande.
- **Modelli Clementine.** I modelli di regressione lineare vengono gestiti come descritto per i modelli SPSS. I modelli di regressione logistica vengono gestiti come descritto per i modelli di regressione logistica. I modelli C&RT Tree vengono gestiti come descritto per i modelli C&RT nell'ambito dei modelli AnswerTree.
- **Modelli di regressione logistica.** Per le covariate dei modelli di regressione logistica, se nel modello salvato è stato incluso un valore medio del predittore, tale valore verrà utilizzato in luogo del valore mancante nel calcolo del punteggio, in modo da consentire la continuazione del calcolo stesso. Se il predittore è categoriale (ad esempio un fattore di un modello di regressione logistica) o se il valore medio non è disponibile, verrà restituito il valore mancante di sistema.

Usa mancante di sistema. Restituisce il valore mancante di sistema durante il calcolo del punteggio in un caso che presenta un valore mancante.

Visualizzazione di un elenco di modelli caricati

È possibile ottenere un elenco dei modelli correntemente caricati. Dai menu (disponibili solo in modalità di analisi distribuita) scegliere:

Trasforma
  Prepara modello
    Elenca modelli

Verrà generata una tabella degli handle di modello, contenente un elenco di tutti i modelli correntemente caricati, incluso il nome (definito anche handle di modello) assegnato al modello, il tipo di modello, il percorso del file modello e il metodo applicato per gestire i valori mancanti.

Figura 8-35
Elenco dei modelli caricati

Model Handles

Name	Type	Input File Name	Validation handling	Missing handling
mregression	NOMREG	C:\models\mregression.xml	ON	SUBSTITUTE
tree	TREE	C:\models\tree.xml	ON	SUBSTITUTE

Funzionalità aggiuntive disponibili con la sintassi dei comandi

Nella finestra di dialogo Carica modello è possibile incollare le proprie selezioni in una finestra di sintassi e modificare la sintassi dei comandi `MODEL HANDLE` risultante. Questo consente di:

- Associare le variabili del modello originale a diverse variabili nel file di dati attivo (utilizzando il sottocomando `MAP`). Per default, il modello viene applicato alle variabili del file di dati attivo che possiedono lo stesso nome delle variabili presenti nel file modello.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo

9

Gestione e trasformazione di file

I file di dati non sempre sono organizzati nella forma ottimale per esigenze specifiche. Può essere necessario combinare i file di dati, ordinare i dati diversamente, selezionare un sottoinsieme di casi oppure modificare l'unità di analisi raggruppando i dati. In SPSS è disponibile una vasta gamma di funzioni per la trasformazione dei dati, tra cui la possibilità di:

Ordinare i dati. È possibile ordinare i dati in base al valore di una o più variabili.

Trasporre casi e variabili. Con il formato di file di dati di SPSS, le righe vengono lette come casi e le colonne come variabili. Per i file di dati in cui tale ordine risulta invertito, è possibile scambiare le righe e le colonne e leggere i dati nel formato corretto.

Unire file. È possibile unire due o più file di dati e inoltre combinare i file con le stesse variabili e casi diversi oppure i file con gli stessi casi e variabili diverse.

Selezionare sottoinsiemi di casi. È possibile limitare l'analisi a un sottoinsieme di casi oppure eseguire contemporaneamente più analisi in diversi sottoinsiemi.

Aggregare i dati. È possibile cambiare l'unità di analisi aggregando i dati in base al valore di una o più variabili di raggruppamento.

Ponderare i dati. Ponderare i dati per l'analisi in base al valore di una variabile di ponderazione.

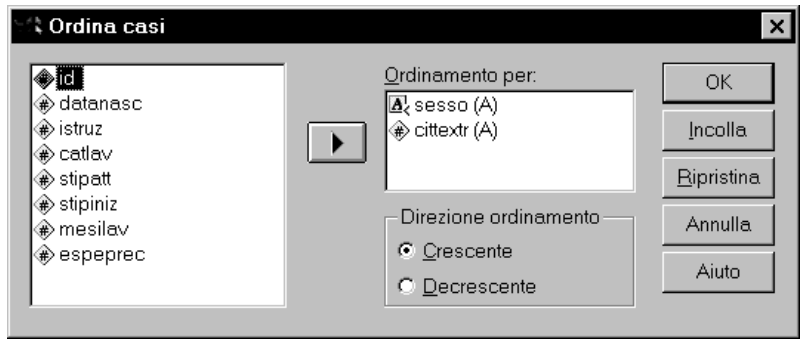
Ristrutturare i dati. È possibile ristrutturare i dati per creare un singolo caso (record) da più casi oppure creare più casi da un singolo caso.

Ordina casi

In questa finestra di dialogo i casi del file di dati vengono ordinati in base ai valori di una o più variabili di ordinamento. È possibile ordinare i casi in ordine crescente o decrescente.

- Se si selezionano più variabili di ordinamento, i casi verranno ordinati in base a ciascuna variabile all'interno delle categorie della variabile precedente nella lista di ordinamento. Se, ad esempio, si seleziona *sesso* come prima variabile di ordinamento e *minoranza* come seconda variabile di ordinamento, i casi verranno classificati per minoranza all'interno di ciascuna categoria di sesso.
- Le variabili stringa verranno ordinate in modo che le lettere maiuscole precedano le minuscole corrispondenti. Ad esempio, la stringa "Sì" precede la stringa "sì" nell'ordinamento.

Figura 9-1
Finestra di dialogo Ordina casi.



Per ordinare i casi

▶ Dai menu, scegliere:

Dati
Ordina casi...

▶ Selezionare una o più variabili di ordinamento.

Trasponi

L'opzione Trasponi consente di creare un nuovo file di dati in cui le righe e le colonne dei file di dati originali vengono trasposte in modo che i casi (righe) diventino variabili (colonne) e viceversa. Verranno creati automaticamente nuovi nomi di variabile che compariranno in una lista specifica.

- Verrà creata automaticamente una nuova variabile stringa, *caso_lbl*, contenente il nome della variabile originale.
- Se il file dati attivo contiene una variabile nome o identificativa con valori univoci, è possibile utilizzarla come Variabile nome. I relativi valori verranno utilizzati come nomi di variabili nel file di dati unito. Se si tratta di una variabile numerica, i nomi di variabile iniziano con la lettera *V* seguita dal valore numerico.
- Nel file di dati trasposto i valori mancanti definiti dall'utente verranno convertiti in valori mancanti di sistema. Se si desidera mantenere i valori mancanti come definiti dall'utente, modificare la definizione di valori mancanti nella Visualizzazione variabili dell'Editor dei dati.

Per trasporre variabili e casi

▶ Dai menu, scegliere:

Dati
  Trasponi...

▶ Selezionare una o più variabili da trasporre in casi.

Unione di file di dati

È possibile unire i dati contenuti in due file in due diversi modi. È possibile:

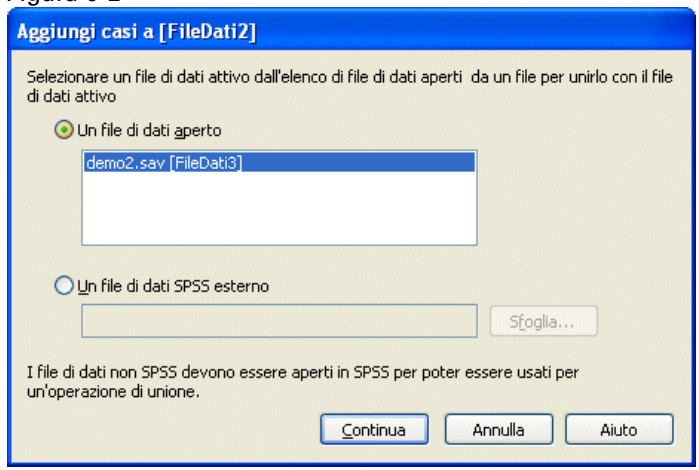
- Unire il file di dati attivo con un altro file di dati aperto oppure con un file di dati in formato SPSS che contenga le stesse variabili ma casi diversi.
- Unire il file di dati attivo con un altro file di dati aperto oppure con un file di dati in formato SPSS che contenga gli stessi casi ma variabili diverse.

Per unire i file:

- Dai menu, scegliere:
  Dati
  Unisci file

- Selezionare Aggiungi casi o Aggiungi variabili.

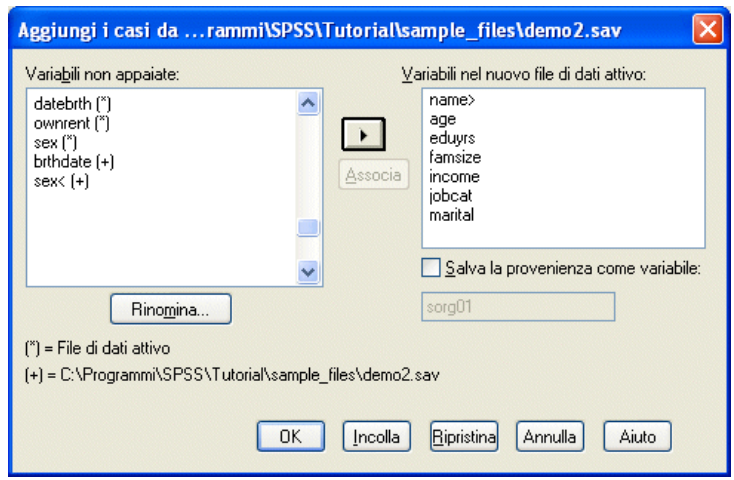
Figura 9-2


Aggiungi casi a [FileDati2]
Selezionare un file di dati attivo dall'elenco di file di dati aperti da un file per unirlo con il file di dati attivo
Un file di dati aperto
demo2.sav [FileDati3]
Un file di dati SPSS esterno
Sfoglia...
I file di dati non SPSS devono essere aperti in SPSS per poter essere usati per un'operazione di unione.
Continua
Annulla
Aiuto

Aggiungi casi

Aggiungi casi unisce il file di dati attivo con un secondo file di dati o un file di dati in formato SPSS che contiene le stesse variabili (colonne) ma casi diversi (righe). È possibile, ad esempio, registrare le stesse informazioni relative ai clienti di due aree di vendita distinte, mantenendo i dati relativi a ciascuna area in un file distinto. Il secondo file di dati può essere un file di dati esterno in formato SPSS o un file di dati della sessione corrente.

Figura 9-3
Finestra di dialogo Aggiungi casi



Variabili non appaiate. Le variabili da escludere dal nuovo file di dati unito. Le variabili del file dati attivo sono identificate dall'asterisco (*). Le variabili di altri file di dati sono identificate dal segno più (+). Per default, la lista contiene:

- Variabili contenute in uno dei due file di dati che non corrispondono a un nome di variabile presente nell'altro file. È possibile creare coppie di variabili non appaiate e quindi includerle nel nuovo file unito.
- Variabili definite come dati numerici in un file e come stringhe nell'altro. Non è possibile unire variabili numeriche con variabili stringa.
- Variabili stringa di lunghezza diversa. La larghezza definita di una variabile stringa deve essere la stessa in entrambi i file di dati.

Variabili in un nuovo file di dati attivo. Le variabili da includere nel nuovo file di dati unito. Per default, tutte le variabili che corrispondono sia al nome che al tipo di dati (numerico o stringa) vengono incluse nella lista.

- È possibile rimuovere dalla lista le variabili che non si desidera includere nel file unito.
- Le variabili non appaiate incluse nel file unito contengono dati mancanti per i casi del file che non contengono la variabile in questione.

Salva la provenienza come variabile. Aggiunge al nuovo file una variabile che assumerà valore 0 per i casi provenienti dal file dati aperto nell'Editor dei dati e valore 1 per i casi provenienti dal file dati esterno.

Per unire file con le stesse variabili e casi diversi

- Aprire almeno uno dei file di dati da unire. Se ci sono più file di dati aperti, impostare uno dei file di dati da unire come file di dati attivo. I casi di questo file appariranno per primi nel nuovo file di dati unito.

- Dai menu, scegliere:
  Dati
  Unisci file
  Aggiungi casi...

- Selezionare il file di dati o il file di dati in formato SPSS da unire con il file di dati attivo.

- Rimuovere tutte le variabili indesiderate dall'elenco Variabili nel nuovo file di dati attivo.

- Aggiungere una coppia di variabili inclusa nella lista Variabili non appaiate che rappresenta le stesse informazioni registrate sotto nomi di variabili diverse nei due file. Alla data di nascita, ad esempio, è possibile che in un file corrisponda il nome di variabile *datanasc* e nell'altro *datanasci*.

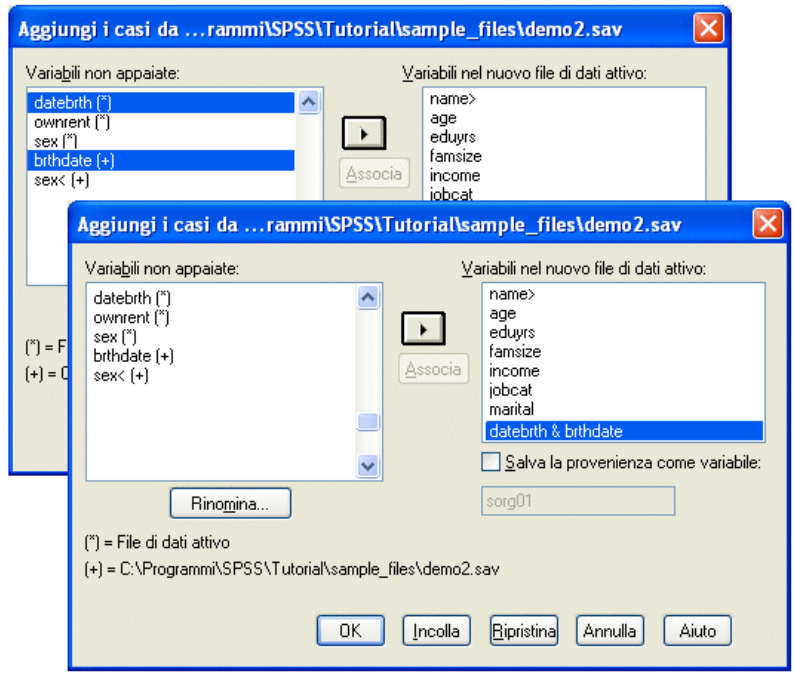
Per selezionare una coppia di variabili non appaiate

- Fare clic su una delle variabili incluse nella lista Variabili non appaiate.

- Fare Ctrl+clic sull'altra variabile nella lista. (Premere Ctrl e contemporaneamente fare clic con il pulsante sinistro del mouse).

- Fare clic su Associa per spostare la coppia di variabili nell'elenco Variabili nel nuovo file dati attivo. Il nome della variabile incluso nel file dati attivo verrà utilizzato come nome di variabile nel file unito.

Figura 9-4
Selezione di coppie di variabili con Ctrl+clic



Aggiungi casi: Rinomina

È possibile rinominare le variabili del file dati attivo o dell'altro file di dati prima di spostarle dalla lista delle variabili non appaiate alla lista delle variabili da includere nel file di dati unito. Rinominando le variabili è possibile:

- Utilizzare un nome di variabile dell'altro file di dati anziché il nome del file dati attivo per le coppie di variabili.
- Includere due variabili con lo stesso nome ma di tipi non appaiati o con stringhe di lunghezza diversa. Ad esempio, per includere sia la variabile numerica *sesso* del file dati attivo che la variabile stringa *sesso* dell'altro file di dati, è innanzitutto necessario rinominare una delle due variabili.

Aggiungi casi: informazioni del dizionario

Le informazioni del dizionario esistenti (etichette delle variabili e dei valori, valori mancanti definiti dall'utente, formati di visualizzazione) nel file dati attivo verranno applicate al file di dati unito.

- Se nel file dati attivo non sono definite informazioni del dizionario per una variabile, verranno utilizzate le informazioni del dizionario del file di dati esterno.
- Se il file dati attivo contiene etichette di valori definite o valori definiti dall'utente per una variabile, verranno ignorate le altre etichette di valori o gli altri valori mancanti definiti dall'utente per tale variabile inclusi nell'altro file di dati.

Unione di più sorgenti dati

La sintassi di comando permette di unire fino a 50 file di dati e/o file. Per ulteriori informazioni, vedere il comando `ADD FILES` in *SPSS Command Syntax Reference* (visualizzabile tramite il menu ?).

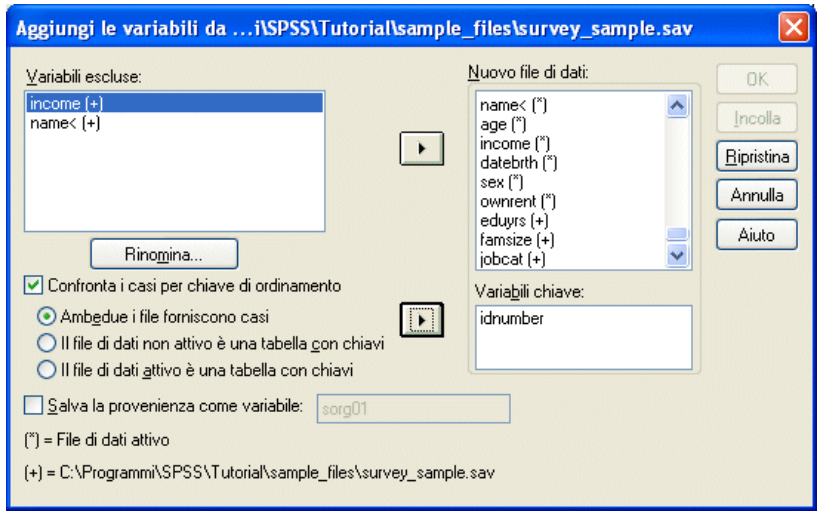
Aggiungi variabili

Aggiungi variabili unisce il file di dati attivo con un altro file di dati o un file di dati in formato SPSS che contiene gli stessi casi (righe) ma variabili diverse (colonne). Ad esempio, può essere necessario unire un file di dati contenente i risultati precedenti a un test con un file contenente i risultati successivi al test.

- I casi devono essere ordinati nello stesso modo in entrambi i file di dati.
- Se per confrontare i casi vengono utilizzate una o più variabili chiave, i due file di dati devono essere disposti nell'ordine crescente definito per tali variabili.
- I nomi delle variabili del secondo file di dati uguali ai nomi delle variabili del file dati attivo verranno esclusi per default in quanto si presume che tali variabili contengano informazioni duplicate.

Salva la provenienza come variabile. Aggiunge al nuovo file una variabile che assumerà valore 0 per i casi provenienti dal file dati aperto nell'Editor dei dati e valore 1 per i casi provenienti dal file dati esterno.

Figura 9-5
Finestra di dialogo Aggiungi variabili



Variabili escluse. Le variabili da escludere dal nuovo file di dati unito. Per default la lista contiene i nomi delle variabili dell'altro file di dati uguali a quelli del file dati attivo. Le variabili del file dati attivo sono identificate dall'asterisco (*). Le variabili di altri file di dati sono identificate dal segno più (+). Se si desidera includere nel file unito una variabile esclusa con un nome duplicato, è possibile rinominarla e aggiungerla alla lista delle variabili da includere.

Nuovo file di dati attivo. Le variabili da includere nel nuovo file di dati unito. Per default, tutti i nomi di variabile univoci in entrambi i file di dati verranno inclusi nella lista.

Variabili chiave. Se per alcuni casi inclusi in un file di dati non esistono file corrispondenti nell'altro file di dati (ovvero, se in uno dei file di dati mancano alcuni casi), utilizzare le variabili chiave per identificare e confrontare correttamente i casi di entrambi i file di dati. È inoltre possibile utilizzare le variabili con i file della tabella di consultazione.

- Le variabili chiave devono avere lo stesso nome in entrambi i file di dati.

- Entrambi i file di dati devono essere disposti nell'ordine crescente delle variabili chiave e l'ordine delle variabili della lista Variabili chiave deve corrispondere alla loro sequenza di ordinamento.
- I casi non corrispondenti delle variabili chiave verranno inclusi nel file unito ma non verranno uniti con i casi dell'altro file. I casi non appaiati contengono solo i valori delle variabili incluse nel file da cui provengono; le variabili dell'altro file contengono il valore mancante di sistema.

Il file di dati non attivo o attivo è una tabella con chiave univoca. La chiave univoca, o **tabella di consultazione**, è un file in cui i dati per ciascun "caso" possono essere applicati a più casi inclusi nell'altro file dati. Ad esempio, se un file contiene informazioni sui singoli membri di una famiglia (ad esempio sesso, età, istruzione) e l'altro file contiene informazioni generali sulla famiglia (ad esempio reddito totale, dimensioni, residenza), è possibile utilizzare il file contenente i dati familiari come tabella di consultazione e applicare i dati comuni sulla famiglia a ciascun membro incluso nel file di dati unito.

Per unire file con gli stessi casi e variabili diverse

▶ Aprire almeno uno dei file di dati da unire. Se ci sono più file di dati aperti, impostare uno dei file di dati da unire come file di dati attivo.

▶ Dai menu, scegliere:

Dati
  Unisci file
    Aggiungi variabili...

▶ Selezionare il file di dati o il file di dati in formato SPSS da unire con il file di dati attivo.

Per selezionare variabili chiave

▶ Selezionare alcune variabili del file esterno (+) nella lista Variabili escluse.

▶ Selezionare Confronta i casi per chiave di ordinamento.

▶ Aggiungere le variabili alla lista Variabili chiave.

Le variabili chiave devono essere presenti sia nel file di dati attivo che nell'altro file di dati. Entrambi i file di dati devono essere disposti nell'ordine crescente delle variabili chiave e l'ordine delle variabili della lista Variabili chiave deve corrispondere alla loro sequenza di ordinamento.

Aggiungi variabili: Rinomina

È possibile rinominare le variabili del file dati attivo o dell'altro file di dati prima di spostarle dalla lista delle variabili alla lista delle variabili da includere nel file di dati unito. Questo risulta estremamente utile se si desidera includere nei due file due variabili con lo stesso nome che contengono informazioni diverse.

Unione di più sorgenti dati

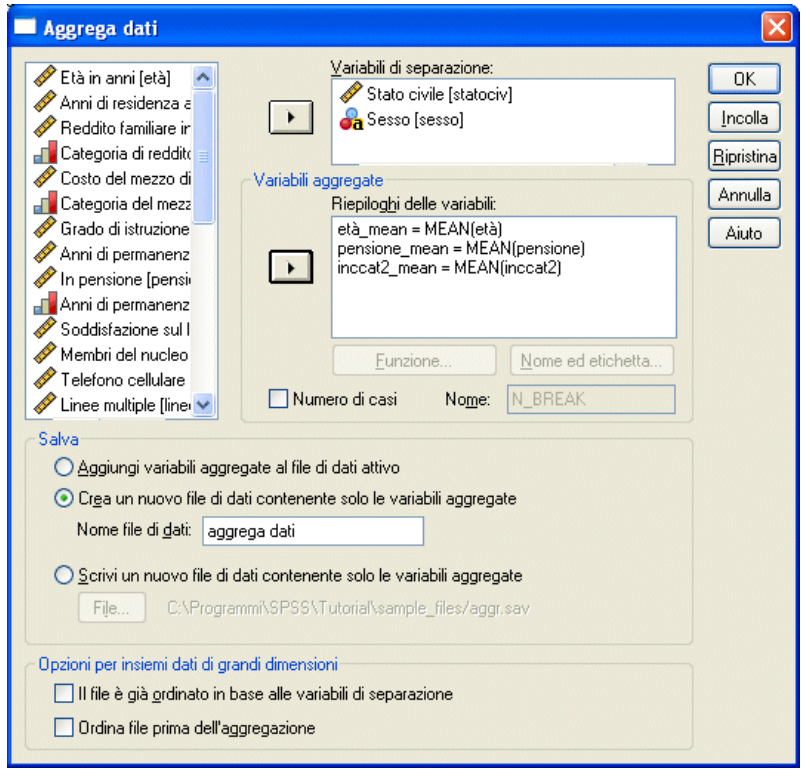
La sintassi di comando permette di unire fino a 50 file di dati e/o file. Per ulteriori informazioni, vedere il comando `MATCH FILES` in *SPSS Command Syntax Reference* visualizzabile tramite il menu ?.

Aggrega dati

Aggrega dati aggrega i gruppi di casi nel file dei dati attivo in singoli casi e creare un nuovo file aggregato oppure nuove variabili nel file dei dati attivo che contengano i dati aggregati. I casi vengono aggregati in base al valore di una o più variabili di separazione (raggruppamento).

- Se si crea un nuovo file di dati aggregato, il nuovo file di dati contiene un caso per ciascun gruppo definito per mezzo delle variabili di separazione. Ad esempio, se la variabile di separazione ha due valori, il nuovo file di dati contiene solo due casi.
- Se si aggiungono variabili aggregate al file di dati attivo, questo non viene aggregato. A ciascun caso con gli stessi valori delle variabili di separazione vengono assegnati gli stessi valori delle nuove variabili aggregate. Ad esempio, se il sesso è l'unica variabile di separazione, a tutti gli uomini viene assegnato lo stesso valore per la nuova variabile aggregata che rappresenta l'età media.

Figura 9-6
Finestra di dialogo Aggrega dati



Variabili di separazione. I casi vengono raggruppati in base ai valori delle variabili di separazione. Ciascuna combinazione univoca di valori delle variabili di separazione costituisce un gruppo. Quando si crea un nuovo file di dati aggregato, tutte le variabili di separazione vengono salvate nel nuovo file con i nomi e le informazioni del dizionario esistenti. La variabile di separazione può essere di tipo numerico o stringa.

Variabili aggregate. Le variabili sorgente vengono utilizzate con le funzioni aggregate per creare nuove variabili aggregate. Il nome della variabile aggregata è seguito da un'etichetta di variabile facoltativa tra virgolette, dal nome della funzione di aggregazione e dal nome della variabile di origine tra parentesi. Le variabili di origine per le funzioni di aggregazione devono essere numeriche.

È possibile annullare i nomi delle variabili aggregate di default indicando nuovi nomi di variabile, fornire etichette di variabile descrittive e modificare le funzioni utilizzate per calcolare i valori aggregati. È inoltre possibile creare una variabile che contenga il numero di casi di ciascun gruppo di interruzione.

Per aggregare un file di dati

- Dai menu, scegliere:
  Dati
  Aggrega...
- Selezionare una o più variabili di separazione che definiscano le modalità di raggruppamento dei casi per creare dati aggregati.
- Selezionare una o più variabili aggregate.
- Selezionare una funzione di aggregazione per ciascuna variabile aggregata.

Salvataggio dei risultati aggregati

È possibile aggiungere variabili aggregate al file di dati attivo oppure creare un nuovo file dati aggregato.

- **Aggiungi le variabili aggregate al file di lavoro.** Al file di lavoro sono aggiunte nuove variabili basate sulle funzioni di aggregazione. Il file di dati non viene aggregato, ma ogni caso con i medesimi valori delle variabili di separazione riceve gli stessi valori per le variabili aggregate.
- **Creare un nuovo file di dati contenente solo le variabili aggregate.** Salva i dati aggregati in un nuovo file di dati nella sessione corrente. Il file di dati include le variabili di raggruppamento che definistono i casi aggregati e tutte le variabili aggregate definite dalle funzioni di aggregazione. Il file di dati attivo non viene modificato.
- **Scrivere un nuovo file di dati contenente solo le variabili aggregate.** Crea un nuovo file contenente i dati aggregati. Il pulsante File consente di cambiare il nome e la collocazione,

Opzioni di ordinamento per i file di dati di grandi dimensioni

Per i file di grandi dimensioni, può essere più conveniente aggregare i dati pre-ordinati.

Il file è già ordinato sulle variabili di separazione. Se i dati sono già stati ordinati sui valori delle variabili di separazione, questa opzione consente di eseguire più velocemente la procedura e di usare meno memoria. Usare questa opzione con cautela.

- I dati devono essere ordinati in base ai valori delle variabili di separazione e nello stesso ordine delle variabili di separazione specificate per la procedura Aggrega dati.
- Se si aggiungono variabili al file dei dati attivo, è necessario selezionare questa opzione solo se i dati vengono ordinati in ordine crescente dei valori delle variabili di separazione.

Ordina il file prima di aggregare. In particolari circostanze con file dati molto grandi, potrebbe essere necessario ordinare il file sui valori delle variabili di separazione prima di aggregare. Questa opzione non è raccomandata, a meno che non si verifichino problemi di memoria o di prestazioni.

Aggrega dati: Funzione di aggregazione

In questa finestra di dialogo è possibile specificare la funzione da utilizzare per calcolare i valori aggregati per le variabili selezionate nella lista Variabili aggregate nella finestra di dialogo Aggrega dati. Di seguito vengono elencate alcune funzioni di aggregazione:

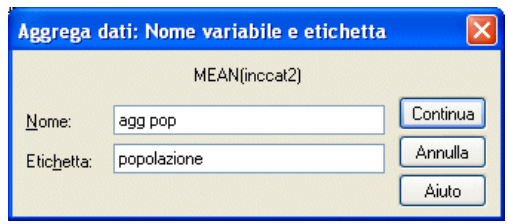
- Funzioni di rappresentazione, ad esempio media, mediana, deviazione standard e somma.
- Numero di casi, compresi i casi pesati, non pesati, mancanti e non mancanti.
- Percentuale o frazioni di valori maggiori o minori del valore specificato.
- Percentuale o frazione di valori interni o esterni all'intervallo specificato.

Figura 9-7
Finestra di dialogo Funzione di aggregazione

Aggrega dati: Nome variabile e etichetta

La funzione Aggrega dati consente di assegnare nomi di variabili di default alle variabili aggregate del nuovo file di dati. In questa finestra di dialogo è possibile modificare il nome della variabile selezionata nella lista Variabili aggregate e specificare un'etichetta descrittiva per la variabile. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.

Figura 9-8
Finestra di dialogo Nome variabile e etichetta

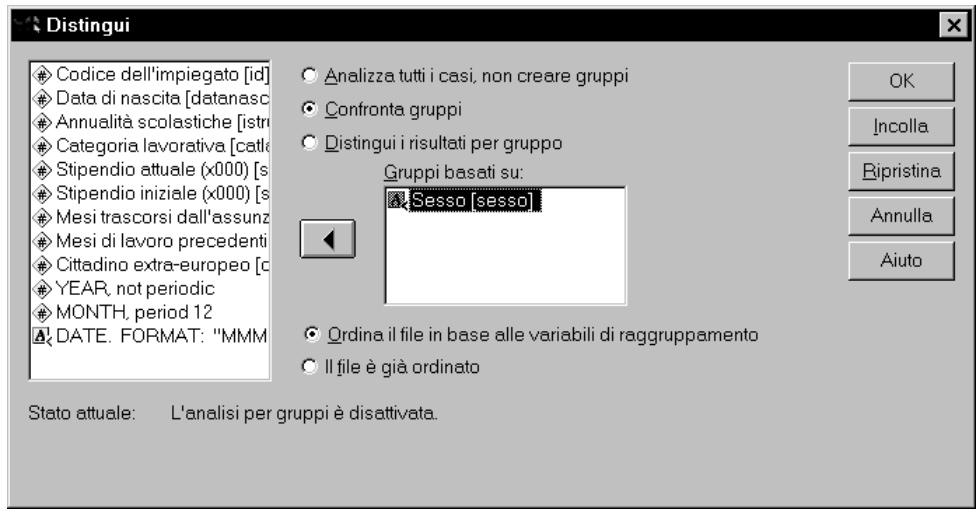


Distingui

L'opzione Distingui consente di suddividere il file di dati in gruppi distinti per l'analisi in base ai valori di una o più variabili di raggruppamento. Se vengono selezionate più variabili di raggruppamento, i casi verranno raggruppati in base a ciascuna variabile all'interno delle categorie della variabile precedente nella lista Gruppi basati su. Se, ad esempio, si seleziona *sesso* come prima variabile di raggruppamento e *minoranza* come seconda variabile di raggruppamento, i casi verranno classificati per minoranza all'interno di ciascuna categoria di sesso.

- È possibile specificare al massimo otto variabili di raggruppamento.
- Ogni gruppo di otto caratteri di una variabile stringa lunga (variabili stringa composte da più di otto caratteri) viene contato come una variabile e anche in questo caso vale il limite massimo di otto variabili di raggruppamento.
- I casi devono essere ordinati in base ai valori della variabile di raggruppamento e nello stesso ordine in cui le variabili compaiono nell'elenco Gruppi basati su. Se il file dati non è già ordinato, selezionare Ordina il file in base alle variabili di raggruppamento.

Figura 9-9
Finestra di dialogo Distingui



Confronta gruppi. Le analisi vengono effettuate sui singoli gruppi, ma i risultati vengono riportati in una sola tabella riepilogativa per facilitare i confronti. Per quanto riguarda le tabelle pivot, viene creata una sola tabella e ciascuna variabile di distinzione può essere spostata tra le dimensioni della tabella. Per quanto riguarda i grafici, viene creato un grafico distinto per ciascun gruppo di analisi; i grafici verranno visualizzati insieme nel Viewer.

Distingui i risultati per gruppo. Tutti i risultati di ciascuna procedura verranno visualizzati separatamente per ciascun gruppo di analisi.

Per suddividere un file di dati in blocchi di dati per l'analisi

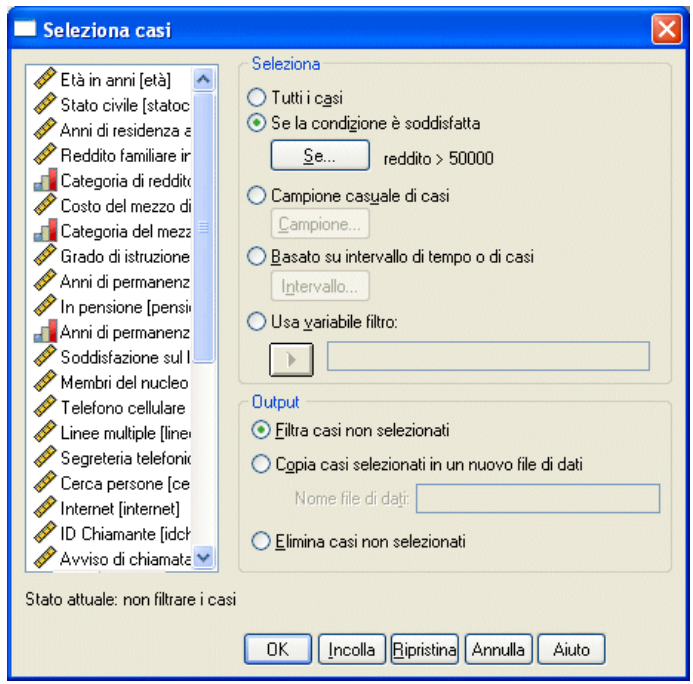
- Dai menu, scegliere:
  Dati
  Distingui...
- Selezionare Confronta gruppi o Distingui i risultati per gruppo.
- Selezionare una o più variabili di raggruppamento.

Seleziona casi

L'opzione Seleziona casi fornisce diversi metodi di selezione di un sottoinsieme di casi in base a criteri che includono variabili ed espressioni complesse. È inoltre possibile selezionare un campione casuale di casi. I criteri in base ai quali viene definito un sottogruppo includono:

- Valori di variabili e intervalli
- Intervalli di date e orari
- Numeri di casi (righe)
- Espressioni aritmetiche
- Espressioni logiche
- Funzioni

Figura 9-10
Finestra di dialogo Seleziona casi



Tutti i casi. Sopprime il filtro corrente e usa tutti i casi nelle analisi successive.

Se la condizione è soddisfatta. Usa un'espressione logica per filtrare i casi. I casi che non soddisfano la condizione impostata vengono esclusi dal campione.

Campione casuale di casi. Seleziona casualmente una percentuale approssimativa o un numero preciso di casi.

Basato su intervallo di tempo o di casi. Seleziona i casi in base ad un intervallo di casi o di date.

Usa variabile filtro. Usa la variabile numerica specificata per selezionare i casi. I casi che assumono valore 0 o mancante per questa variabile vengono esclusi dal campione.

Uscita

Questa sezione controlla il trattamento dei casi non selezionati. Per il trattamento dei casi non selezionati sono disponibili le opzioni seguenti:

- **Rimuovi casi non selezionati.** I casi non selezionati non saranno inclusi nell'analisi, ma rimarranno nel file di dati. e sarà possibile utilizzarli successivamente nella sessione se si disattivano le funzioni di filtro. Se si seleziona un campione casuale o si selezionano casi in base a un'espressione logica, verrà creata una variabile denominata *filtro_$* con un valore 1 per i casi selezionati e un valore 0 per i casi non selezionati.
- **Copia casi selezionati in un nuovo file di dati.** I casi selezionati vengono copiati in un nuovo file di dati senza modificare quello originale. I casi non selezionati non vengono inclusi nel nuovo file di lavoro e vengono lasciati nello stato originale nel file di dati originale.
- **Elimina casi non selezionati.** I casi non selezionati verranno cancellati dal file di dati. Per recuperare i casi cancellati è necessario uscire dal file senza salvare le modifiche e quindi riaprire il file. Se si salvano le modifiche al file di dati, l'eliminazione dei casi sarà definitiva. *Nota*: Se si eliminano i casi non selezionati e si salva il file, non sarà possibile recuperare i casi.

Per selezionare un sottoinsieme di casi

▶ Dai menu, scegliere:
Dati
Seleziona casi...

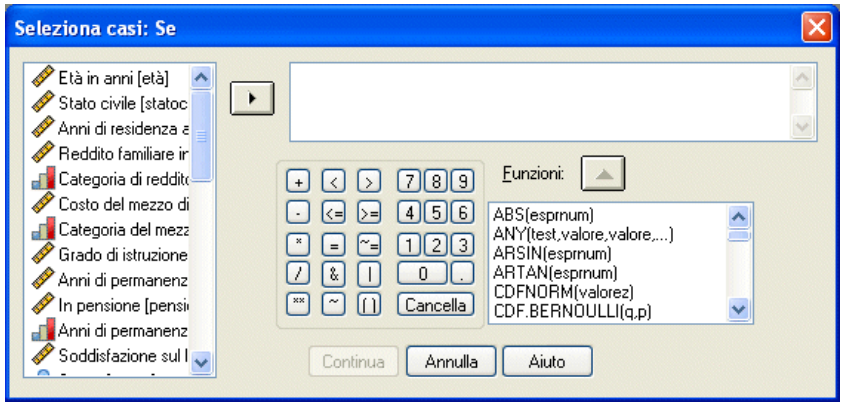
▶ Selezionare uno dei metodi di selezione dei casi.

▶ Specificare i criteri di selezione dei casi.

Seleziona casi: Se

In questa finestra di dialogo è possibile selezionare sottoinsiemi di casi utilizzando espressioni logiche. Un'espressione logica restituisce un valore *vero*, *falso*, o *mancante* per ciascun caso.

Figura 9-11
Finestra di dialogo Seleziona casi: Se



- Se il risultato di un'espressione logica è *vero*, il caso verrà incluso nel sottoinsieme selezionato.
- Se il risultato di un'espressione logica è *falso* o *mancante*, il caso non viene incluso nel sottoinsieme selezionato.
- La maggior parte delle espressioni condizionali utilizza uno o più dei sei operatori relazionali (<, >, <=, >=, = e ~=) della calcolatrice.
- Le espressioni logiche possono includere nomi di variabili, costanti, operatori aritmetici, funzioni numeriche e di altro tipo, variabili logiche e operatori relazionali.

Seleziona casi: Campione casuale

In questa finestra di dialogo è possibile selezionare un campione casuale in base a una percentuale approssimativa o a un numero esatto di casi. Il campionamento viene eseguito senza sostituzione: in tal modo è impossibile selezionare lo stesso caso più di una volta.

Figura 9-12
Finestra di dialogo Seleziona casi: Campione casuale

Approssimativamente. Genera un campione casuale che include approssimativamente la percentuale di casi specificata. Poiché per ciascun caso viene eseguito un processo indipendente di decisione pseudo-casuale, l'equivalenza tra la percentuale di casi selezionati e la percentuale specificata può essere solo approssimativa. Maggiore è il numero di casi inclusi nel data file e maggiore sarà l'approssimazione della percentuale di casi selezionati rispetto alla percentuale specificata.

Esattamente. Un numero di casi specificato dall'utente. È necessario specificare anche il numero di casi in base al quale generare il campione. Il secondo numero deve essere uguale o minore del numero totale di casi del file dati. Se questo numero supera il numero totale di casi del file di dati, in proporzione il campione conterrà un numero di dati inferiore a quello specificato.

Seleziona casi: Intervallo

In questa finestra di dialogo è possibile selezionare casi in base a un intervallo di numeri di casi o a un intervallo di date e orari.

- Gli intervalli di casi si basano sul numero di riga visualizzato nell'Editor dei dati.
- Gli intervalli di date e orari sono disponibili solo per i **dati di serie storiche** con variabili della data definite (menu Dati, Definisci date).

Figura 9-13
Finestra di dialogo Seleziona casi: intervallo per intervalli di casi (nessuna variabile di data definita)

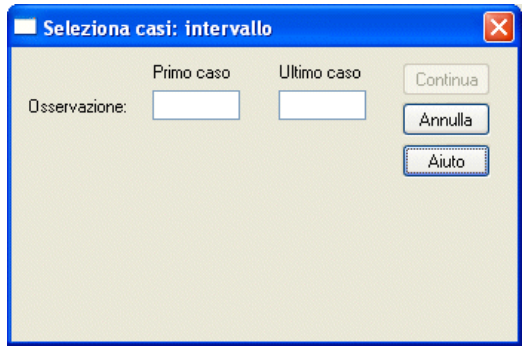
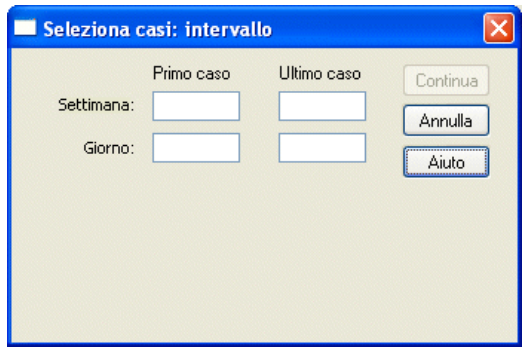


Figura 9-14
Finestra di dialogo Seleziona casi: intervallo per dati di serie storiche con variabili di data definite



Pesa casi

L'opzione Pesa casi consente di assegnare ai casi pesi diversi (mediante replicazione simulata) per l'analisi statistica.

- I valori della variabile di ponderazione devono indicare il numero di osservazioni rappresentate dai singoli casi del file di dati.

- I casi con valori uguali a zero, negativi o mancanti per la variabile di ponderazione verranno esclusi dall'analisi.
- I valori frazionari sono validi e vengono utilizzati solo quando è opportuno, in genere quando i casi vengono riportati in tabelle.

Figura 9-15
Finestra di dialogo Pesa casi

Una volta applicata, una variabile di ponderazione rimane valida finché non ne viene selezionata un'altra o si disattiva la ponderazione. Quando si salva un file di dati ponderato, le informazioni di ponderazione verranno salvate insieme al file. È possibile disattivare la ponderazione in qualsiasi momento, anche dopo aver salvato il file in forma ponderata.

Ponderazione nelle tavole di contingenza. Nella procedura Tavole di contingenza sono disponibili diverse opzioni per la gestione dei pesi dei casi. Per ulteriori informazioni, vedere "Tavole di contingenza: Visualizzazione cella" in Capitolo 17 a pag. 376.

Ponderazione in grafici a dispersione e istogrammi. Per i grafici a dispersione e gli istogrammi è disponibile un'opzione per attivare e disattivare i pesi dei casi. Tale opzione, tuttavia, non produce alcun effetto sui casi con valore negativo, uguale a 0 o mancante per la variabile di ponderazione. Tali casi verranno esclusi dal grafico anche se la ponderazione viene disattivata dal suo interno.

Per pesare i casi

▶ Dai menu, scegliere:

Dati
  Pesa casi...

▶ Selezionare Pesa i casi per.

▶ Selezionare una variabile di frequenza.

I valori della variabile di frequenza vengono utilizzati come pesi dei casi. Ad esempio un caso con un valore 3 per la variabile di frequenza rappresenterà 3 casi nel file di dati ponderato.

Ristrutturazione di dati

La Ristrutturazione di dati guidata consente di ristrutturare i dati per la procedura SPSS che si desidera utilizzare. La procedura guidata sostituisce il file corrente con un nuovo file ristrutturato e offre le funzioni seguenti:

- Ristruttura le variabili selezionate in casi.
- Ristruttura i casi selezionati in variabili.
- Trasponi tutti i dati.

Per ristrutturare i dati

▶ Dai menu, scegliere:
Dati
Ristruttura…

▶ Selezionare il tipo di ristrutturazione che si desidera eseguire.

▶ Selezionare i dati da ristrutturare.

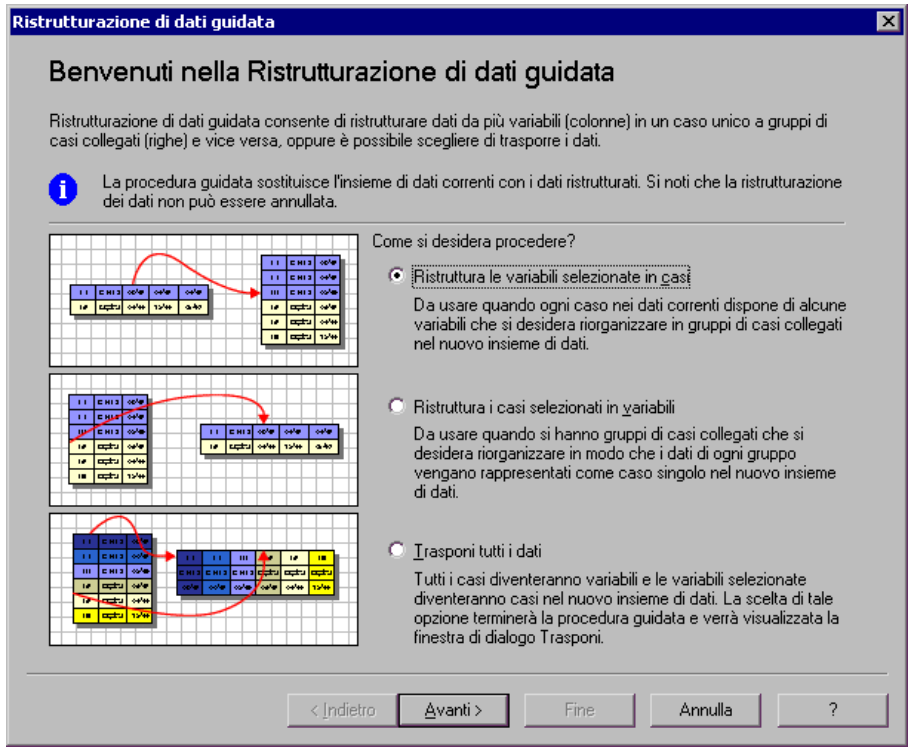
Se lo si desidera, è possibile:

- Creare variabili di identificazione, che consentono di risalire da un valore nel nuovo file a un valore nel file originale.
- Ordinare i dati prima della ristrutturazione.
- Definire le opzioni per il nuovo file.
- Incollare la sintassi dei comandi in una finestra di sintassi.

Ristrutturazione di dati guidata: Seleziona tipo

La Ristrutturazione di dati guidata consente di ristrutturare i dati in uso. Nella prima finestra di dialogo selezionare il tipo di ristrutturazione che si desidera eseguire.

Figura 9-16
Ristrutturazione di dati guidata


Ristrutturazione di dati guidata
Benvenuti nella Ristrutturazione di dati guidata
Ristrutturazione di dati guidata consente di ristrutturare dati da più variabili (colonne) in un caso unico a gruppi di casi collegati (righe) e vice versa, oppure è possibile scegliere di trasporre i dati.
La procedura guidata sostituisce l'insieme di dati correnti con i dati ristrutturati. Si noti che la ristrutturazione dei dati non può essere annullata.
Come si desidera procedere?
Ristruttura le variabili selezionate in casi
Da usare quando ogni caso nei dati correnti dispone di alcune variabili che si desidera riorganizzare in gruppi di casi collegati nel nuovo insieme di dati.
Ristruttura i casi selezionati in variabili
Da usare quando si hanno gruppi di casi collegati che si desidera riorganizzare in modo che i dati di ogni gruppo vengano rappresentati come caso singolo nel nuovo insieme di dati.
Trasponi tutti i dati
Tutti i casi diventeranno variabili e le variabili selezionate diventeranno casi nel nuovo insieme di dati. La scelta di tale opzione terminerà la procedura guidata e verrà visualizzata la finestra di dialogo Trasponi.
< Indietro
Avanti >
Fine
Annulla
?

- **Ristruttura le variabili selezionate in casi.** Selezionare questa opzione se nei dati sono presenti gruppi di colonne correlate che si desidera visualizzare come gruppi di righe nel nuovo file di dati. Se si sceglie questa opzione, verranno visualizzati i passi per **Variabili in casi**.
- **Ristruttura i casi selezionati in variabili.** Selezionare questa opzione se nei dati sono presenti gruppi di righe correlate che si desidera visualizzare come gruppi di colonne nel nuovo file di dati. Se si sceglie questa opzione, verranno visualizzati i passi per **Casi in variabili**.
- **Trasponi tutti i dati.** Selezionare questa opzione se si desidera trasporre i dati. Tutte le righe diventeranno colonne e tutte le colonne diventeranno righe nei nuovi dati. Se si sceglie questa opzione, la Ristrutturazione di dati guidata verrà chiusa e verrà aperta la finestra di dialogo Trasponi dati.

Scelta della modalità di ristrutturazione dei dati

Una **variabile** contiene le informazioni che si desidera analizzare, ad esempio una misura o un punteggio. Un **caso** è un'osservazione, ad esempio un individuo. In una struttura di dati *semplice* ogni variabile è una singola colonna di dati e ogni caso è una singola riga. Se, ad esempio, si misurano i punteggi dei test relativi a tutti gli studenti di una classe, tutti i valori vengono visualizzati in una singola colonna e per ogni studente è presente una singola riga.

L'analisi dei dati comporta in genere l'analisi della modalità di variazione di una variabile in base a una condizione specifica, che può essere rappresentata da una cura sperimentale specifica, da informazioni demografiche, da un punto temporale o da altro. Nell'analisi dei dati, le condizioni prese in considerazione vengono in genere denominate **fattori**. Se si analizzano i fattori, la struttura dei dati sarà *complessa*. Le informazioni relative a una variabile possono essere presenti in più colonne dei dati, ad esempio in una colonna per ogni livello di un fattore, oppure le informazioni relative a un caso possono essere presenti in più righe, ad esempio una riga per ogni livello di un fattore. La Ristrutturazione di dati guidata semplifica la ristrutturazione dei file con una struttura di dati complessa.

Le scelte eseguite nella procedura guidata variano in base alla struttura del file corrente e alla struttura che si desidera creare nel nuovo file.

Disposizione dei dati nel file corrente. I dati correnti possono essere disposti in modo tale che i fattori vengano registrati in una variabile *distinta* (in gruppi di casi) o *con* la variabile (in gruppi di variabili).

- **Gruppi di casi.** Si supponga che le variabili e le condizioni del file corrente siano registrate in colonne distinte. Ad esempio:

var	fattore
8	1
9	1
3	2
1	2

In questo esempio le prime due righe sono un **gruppo di casi** perché sono correlate e contengono i dati relativi allo stesso livello di fattore. Nell'analisi dei dati SPSS, il fattore viene in genere definito **variabile di raggruppamento** quando i dati sono strutturati in questo modo.

- **Gruppi di colonne.** Si supponga che le variabili e le condizioni del file corrente siano registrate nella stessa colonna. Ad esempio:

var_1	var_2
8	3
9	1

In questo esempio, le due colonne sono un **gruppo di variabili** perché sono correlate. Contengono i dati della stessa variabile, *var_1* per il livello di fattore 1 e *var_2* per il livello di fattore 2. Nell'analisi dei dati SPSS il fattore viene in genere definito **misura ripetuta** quando i dati sono strutturati in questo modo.

Disposizione dei dati nel nuovo file. È in genere determinata dalla procedura che si desidera utilizzare per l'analisi dei dati.

- **Procedure che richiedono gruppi di casi.** È necessario che i dati siano strutturati in gruppi di casi se si desidera eseguire analisi che richiedono una variabile di raggruppamento, ad esempio *univariata*, *multivariata* e *componenti della varianza* con la procedura Modello lineare generalizzato, Modelli misti, Cubi OLAP e *campioni indipendenti* con le procedure Test T o Test non parametrici. Se i dati correnti sono strutturati in gruppi di variabili e si desidera eseguire le analisi, selezionare Ristruttura le variabili selezionate in casi.
- **Procedure che richiedono gruppi di variabili.** È necessario che i dati siano strutturati in gruppi di variabili se si desidera analizzare le misure ripetute, ad esempio *misure ripetute* con la procedura Modello lineare generalizzato, l'analisi *covariata dipendente dal tempo* con l'analisi Regressione di Cox, *campioni appaiati* con Test T o *campioni dipendenti* con Test non parametrici. Se i dati correnti sono strutturati in gruppi di casi e si desidera eseguire le analisi, selezionare Ristruttura i casi selezionati in variabili.

Esempio di Variabili in casi

In questo esempio i punteggi dei test vengono registrati in colonne distinte per ogni fattore, *A* e *B*.

Figura 9-17
Dati correnti per Variabili in casi

	punteg_a	punteg_b
1	1014.00	864.00
2	684.00	636.00
3	810.00	638.00

Si desidera eseguire un test *t* per campioni indipendenti ed è disponibile un gruppo di colonne che include *punteggio_a* e *punteggio_b*, ma non è disponibile la **variabile di raggruppamento** richiesta dalla procedura. Selezionare Ristruttura le variabili selezionate in casi nella Ristrutturazione di dati guidata, ristrutturare un gruppo di variabili in una nuova variabile denominata *punteggio* e creare un indice denominato *gruppo*. Il nuovo file di dati è illustrato nella figura seguente.

Figura 9-18
Nuovi dati ristrutturati per Variabili in casi

	grouppo	punteggio
1	PUNTEG_A	1014.00
2	PUNTEG_B	864.00
3	PUNTEG_A	684.00
4	PUNTEG_B	636.00
5	PUNTEG_A	810.00
6	PUNTEG_B	638.00

Quando si esegue il test *t* per campioni indipendenti sarà possibile utilizzare *gruppo* come variabile di raggruppamento.

Esempio di Casi in variabili

In questo esempio i punteggi dei test vengono registrati due volte per ogni soggetto, —prima e dopo il trattamento.

Figura 9-19
Dati correnti per Casi in variabili

	id	punteg	tempo
1	1	1014.00	prec
2	1	864.00	succ
3	2	684.00	prec
4	2	636.00	succ

Si desidera eseguire un test *t* per campioni appaiati. e i dati sono strutturati in gruppi di casi, ma non sono disponibili le **misure ripetute** per le coppie di variabili richieste dalla procedura. Selezionare Ristruttura i casi selezionati in variabili nella Ristrutturazione di dati guidata, utilizzare *id* per identificare i gruppi di righe nei dati correnti e *orario* per creare il gruppo di variabili nel nuovo file.

Figura 9-20
Nuovi dati ristrutturati per Casi in variabili

	id	succ	prec
1	1	864.00	1014.00
2	2	636.00	684.00

Nella procedura Test *t* per campioni appaiati, è possibile utilizzare *prima* e *dopo* come coppia di variabili.

Ristrutturazione di dati guidata (Variabili in casi): Numero di gruppi di variabili

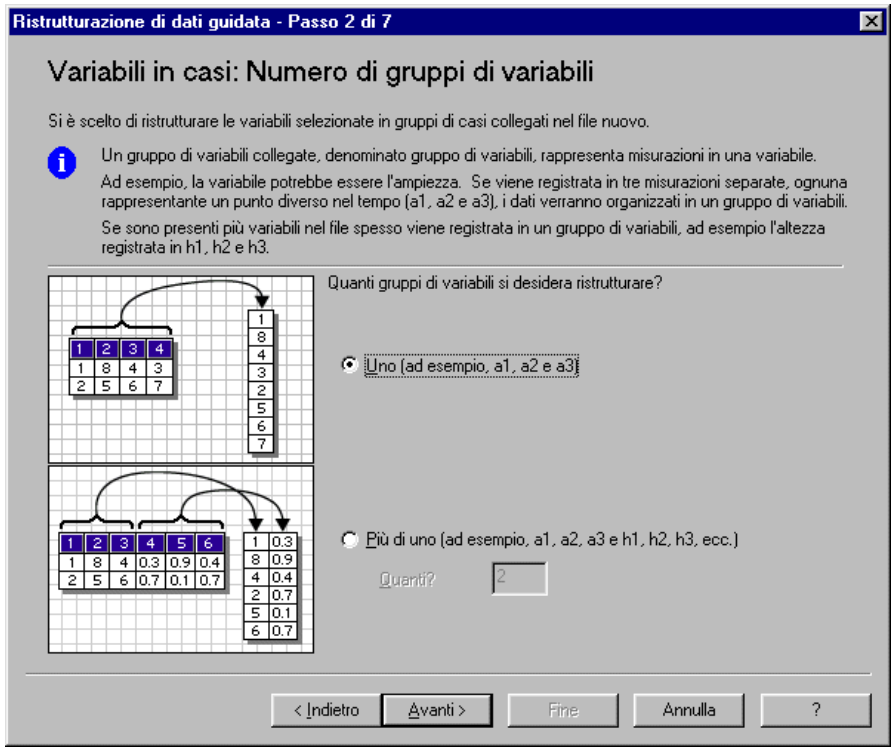
Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe

e consente di scegliere il numero di gruppi di variabili del file corrente che si desidera ristrutturare nel nuovo file.

Numero dei gruppi di variabili nel file corrente. Stabilire quanti gruppi di variabili sono presenti nei dati correnti. Un gruppo di colonne correlate, denominato **gruppo di variabili**, registra le misure ripetute della stessa variabile in colonne distinte. Ad esempio, se nei dati correnti sono presenti le tre colonne *a1*, *a2* e *a3* che registrano **ampiezza**, significa che è presente un gruppo di variabili. Se sono presenti le tre colonne aggiuntive *h1*, *h2* e *h3* che registrano **altezza**, significa che sono disponibili due gruppi di variabili.

Numero dei gruppi di variabili nel nuovo file. Stabilire il numero dei gruppi di variabili che si desidera rappresentare nel nuovo file di dati, tenendo presente che non è necessario ristrutturare tutti i gruppi di variabili nel nuovo file.

Figura 9-21
Ristrutturazione di dati guidata: Numero di gruppi di variabili, Passo 2


Ristrutturazione di dati guidata - Passo 2 di 7
Variabili in casi: Numero di gruppi di variabili
Si è scelto di ristrutturare le variabili selezionate in gruppi di casi collegati nel file nuovo.
Un gruppo di variabili collegate, denominato gruppo di variabili, rappresenta misurazioni in una variabile.
Ad esempio, la variabile potrebbe essere l'ampiezza. Se viene registrata in tre misurazioni separate, ognuna rappresentante un punto diverso nel tempo (a1, a2 e a3), i dati verranno organizzati in un gruppo di variabili.
Se sono presenti più variabili nel file spesso viene registrata in un gruppo di variabili, ad esempio l'altezza registrata in h1, h2 e h3.
Quanti gruppi di variabili si desidera ristrutturare?
Uno (ad esempio, a1, a2 e a3)
Più di uno (ad esempio, a1, a2, a3 e h1, h2, h3, ecc.)
Quanti?
2
< Indietro
Avanti >
Fine
Annulla
?

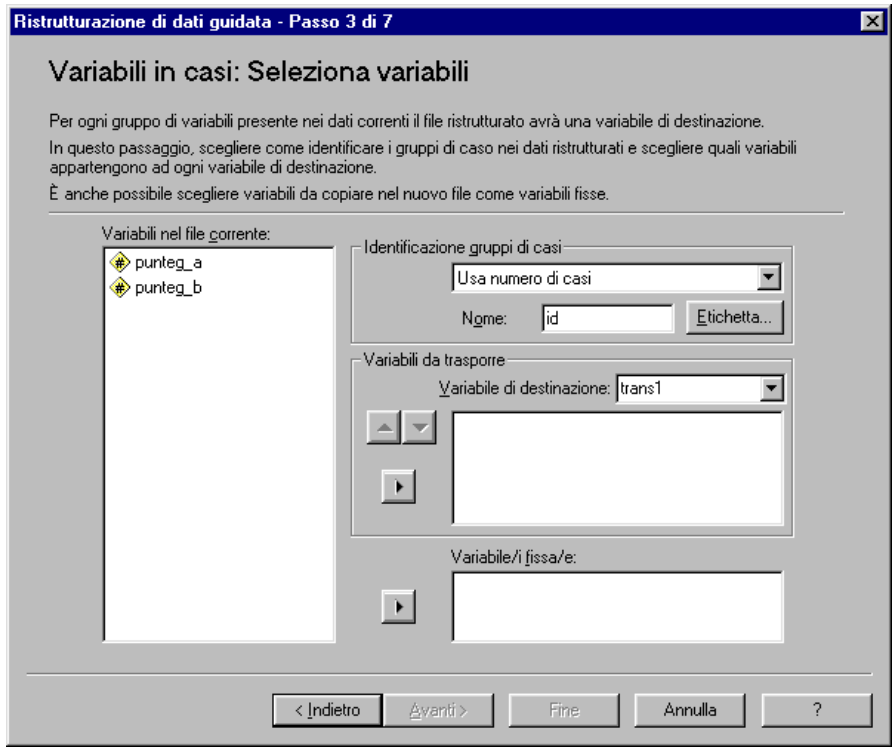
- **Uno.** Nel nuovo file verrà creata una singola variabile ristrutturata in base a un gruppo di variabili del file corrente.
- **Più di uno.** Nel nuovo file verranno create più variabili ristrutturate. Il numero specificato influisce sul passo successivo, nel quale viene creato automaticamente il numero specificato di nuove variabili.

Ristrutturazione di dati guidata (Variabili in casi): Selezionare Variabili

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe

e fornisce informazioni su come utilizzare le variabili del file corrente nel nuovo file. È inoltre possibile creare una variabile che identifica le righe nel nuovo file.

Figura 9-22
Ristrutturazione di dati guidata: Selezione delle variabili, Passo 3


Ristrutturazione di dati guidata - Passo 3 di 7
Variabili in casi: Seleziona variabili
Per ogni gruppo di variabili presente nei dati correnti il file ristrutturato avrà una variabile di destinazione.
In questo passaggio, scegliere come identificare i gruppi di caso nei dati ristrutturati e scegliere quali variabili appartengono ad ogni variabile di destinazione.
È anche possibile scegliere variabili da copiare nel nuovo file come variabili fisse.
Variabili nel file corrente:
punteg_a
punteg_b
Identificazione gruppi di casi
Usa numero di casi
Nome:
id
Etichetta...
Variabili da trasporre
Variabile di destinazione:
trans1
Variabile/i fissa/e:
< Indietro
Avanti >
Fine
Annulla
?

Modalità di identificazione delle nuove righe. È possibile creare nel nuovo file di dati una variabile che identifica la riga del file corrente utilizzata per creare un gruppo di nuove righe. L'identificatore può essere rappresentato da un numero sequenziale di caso oppure dai valori della variabile. Per definire la variabile di identificazione nel nuovo file, utilizzare i controlli disponibili in Identificazione gruppi di casi. Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per la variabile di identificazione.

Elementi da ristrutturare nel nuovo file. Nel passo precedente è stato specificato il numero dei gruppi di variabili che si desidera ristrutturare. Per ogni gruppo è stata creata una nuova variabile e i valori per il gruppo di variabili verranno visualizzati in tale variabile all'interno del nuovo file. Per definire la variabile ristrutturata nel nuovo file, utilizzare i controlli disponibili in Variabili da trasporre.

Per specificare una variabile ristrutturata

- Inserire le variabili che costituiscono il gruppo di variabili che si desidera trasformare nell'elenco Variabili da trasporre. Tutte le variabili nel gruppo devono essere dello stesso tipo (numerico o di stringa).

  È possibile includere la stessa variabile più di una volta nel gruppo di variabili (le variabili vengono copiate anzichè spostate dall'elenco sorgente delle variabili) e i corrispondenti valori verranno ripetuti nel nuovo file.

Per specificare più variabili ristrutturate

- Selezionare la prima variabile di destinazione che si desidera definire dall'elenco a discesa.

- Inserire le variabili che costituiscono il gruppo di variabili che si desidera trasformare nell'elenco Variabili da trasporre. Tutte le variabili nel gruppo devono essere dello stesso tipo (numerico o di stringa). Nel gruppo di variabili è possibile includere la stessa variabile più volte. Una variabile viene copiata invece che spostata dall'elenco di variabili di origine e i relativi valori vengono ripetuti nel nuovo file.

- Selezionare la successiva variabile di destinazione che si desidera definire e ripetere il processo di selezione della variabile per tutte le variabili di destinazione disponibili.
  - Benché sia possibile includere la stessa variabile più di una volta nello stesso gruppo di variabili di destinazione, non è possibile includere la stessa variabile in più di un gruppo di variabili di destinazione.
  - Ogni elenco relativo al gruppo di variabili di destinazione deve contenere lo stesso numero di variabili. Le variabili elencate più di una volta vengono incluse nel conteggio.
  - Il numero di gruppi di variabili di destinazione è determinato dal numero di gruppi di variabili specificati nel passo precedente. Qui è possibile modificare i nomi predefiniti delle variabili, ma è necessario ritornare al passo precedente per modificare il numero dei gruppi di variabili da ristrutturare.
  - È necessario definire i gruppi di variabili (selezionando le variabili dall'elenco sorgente) per tutte le variabili di destinazione disponibili prima di procedere al passo successivo.

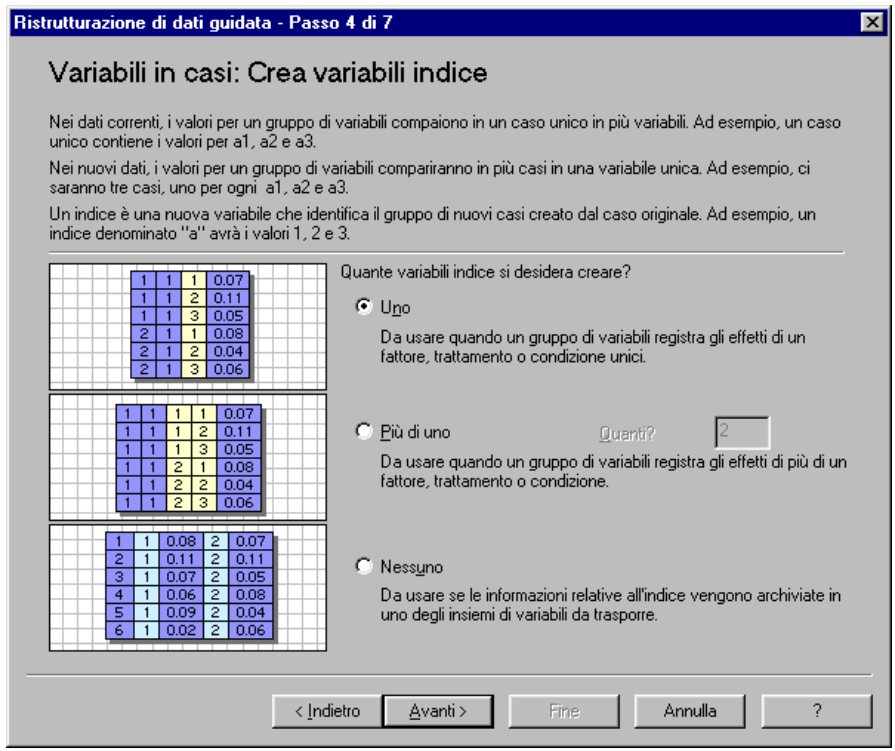
Elementi da copiare nel nuovo file. È possibile copiare nel nuovo file le variabili che non vengono ristrutturate, i cui valori verranno propagati nelle nuove righe. Spostare le variabili che si desidera copiare nel nuovo file nell'elenco Variabile/i fissa/e.

Ristrutturazione di dati guidata (Variabili in casi): Crea variabili indice

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe

e consente di scegliere se creare variabili indice. Un indice è una nuova variabile che identifica in modo sequenziale un gruppo di righe in base alla variabile originale per cui è stata creata la nuova riga.

Figura 9-23
Ristrutturazione di dati guidata: Creazione delle variabili indice, Passo 4


Ristrutturazione di dati guidata - Passo 4 di 7
Variabili in casi: Crea variabili indice
Nei dati correnti, i valori per un gruppo di variabili compaiono in un caso unico in più variabili. Ad esempio, un caso unico contiene i valori per a1, a2 e a3.
Nei nuovi dati, i valori per un gruppo di variabili compariranno in più casi in una variabile unica. Ad esempio, ci saranno tre casi, uno per ogni a1, a2 e a3.
Un indice è una nuova variabile che identifica il gruppo di nuovi casi creato dal caso originale. Ad esempio, un indice denominato "a" avrà i valori 1, 2 e 3.
Quante variabili indice si desidera creare?
Uno
Da usare quando un gruppo di variabili registra gli effetti di un fattore, trattamento o condizione unici.
Più di uno
Quanti?
2
Da usare quando un gruppo di variabili registra gli effetti di più di un fattore, trattamento o condizione.
Nessuno
Da usare se le informazioni relative all'indice vengono archiviate in uno degli insiemi di variabili da trasporre.
< Indietro
Avanti >
Fine
Annulla
?

Numero delle variabili indice nel nuovo file. Le variabili indice possono essere utilizzate come variabili di raggruppamento nelle procedure di SPSS. Nella maggior parte dei casi è sufficiente una singola variabile indice. Se tuttavia i gruppi di variabili nel file corrente rappresentano più livelli di fattore, può essere opportuno prevedere più indici.

- **Uno.** Verrà creata una singola variabile indice.
- **Più di uno.** Verranno creati più indici e sarà possibile specificare il numero degli indici che si desidera creare. Il numero specificato influisce sul passo successivo, nel quale viene creato automaticamente il numero di indici specificato.
- **Nessuno.** Selezionare questa opzione se non si desidera creare variabili indice nel nuovo file.

Esempio di indice singolo per Variabili in casi

Nei dati correnti sono presenti il gruppo di variabili *ampiezza* e il fattore *orario*. L'ampiezza è stata misurata tre volte e registrata in *a1*, *a2* e *a3*.

Figura 9-24
Dati correnti per un Indice

	soggetto	a1	a2	a3
1	1	6.70	4.30	5.70
2	2	7.10	5.90	5.60

Il gruppo di variabili verrà ristrutturato in una singola variabile, *ampiezza*, e verrà creato un singolo indice numerico. I nuovi dati sono illustrati nella tabella seguente.

Figura 9-25
Nuovi dati ristrutturati con un indice

	soggetto	indice	ampiezza
1	1	1	6.70
2	1	2	4.30
3	1	3	5.70
4	2	1	7.10
5	2	2	5.90
6	2	3	5.60

Indice inizia con 1 e viene incrementato per ogni variabile presente nel gruppo. Si riavvia ogni volta che trova una nuova riga nel file originale. Ora è possibile usare *indice* nelle procedure SPSS che richiedono una variabile di raggruppamento.

Esempio di due indici per Variabili in casi

Se un gruppo di variabili registra più fattori, è possibile creare più indici. I dati correnti devono tuttavia essere disposti in modo tale che i livelli del primo fattore rappresentino un indice primario all'interno del quale i livelli dei fattori successivi eseguono dei cicli. Nei dati correnti è presente il gruppo di variabili *ampiezza* e i due fattori *A* e *B*. I dati sono disposti in modo tale che i livelli del fattore *B* eseguano cicli all'interno dei livelli del fattore *A*.

Figura 9-26
Dati correnti per due indici

	soggetto	a_a1b1	a_a1b2	a_a2b1	a_a2b2
1	1	5.50	6.40	5.80	5.90
2	2	7.40	7.10	5.60	6.70

Il gruppo di variabili verrà ristrutturato in una singola variabile, *ampiezza*, e verranno creati due indici. I nuovi dati sono illustrati nella tabella seguente.

Figura 9-27
Nuovi dati ristrutturati con due indici

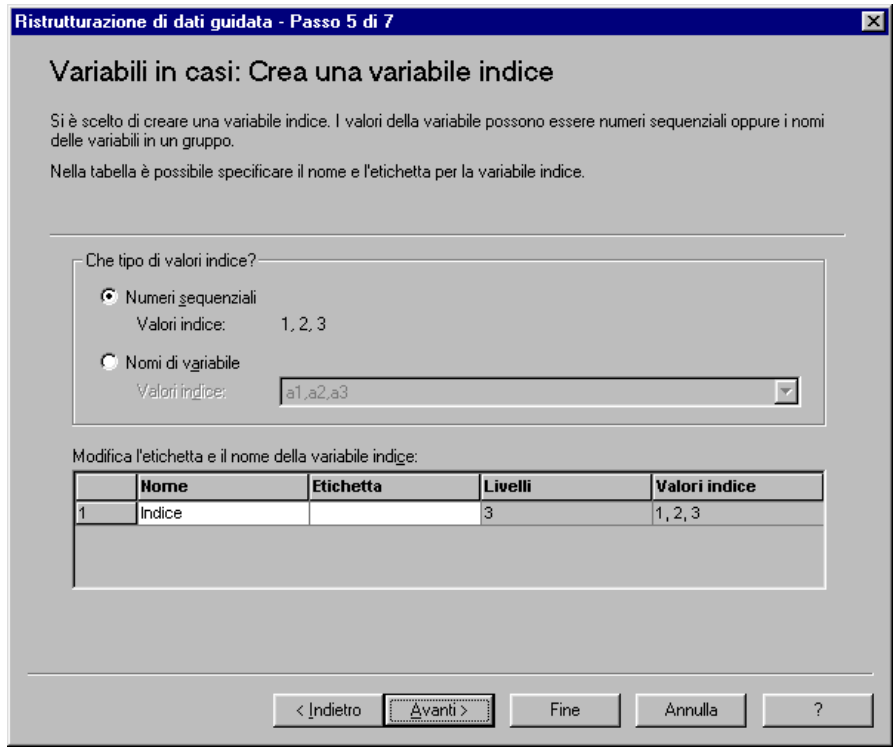
	soggetto	indice_a	indice_b	ampiezza
1	1	1	1	5.50
2	1	1	2	6.40
3	1	2	1	5.80
4	1	2	2	5.90
5	2	1	1	7.40
6	2	1	2	7.10
7	2	2	1	5.60
8	2	2	2	6.70

Ristrutturazione di dati guidata (Variabili in casi): Crea una variabile indice

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e di creare una variabile indice

e consente di stabilire i valori desiderati per la variabile indice. I valori possono essere rappresentati da numeri sequenziali o dai nomi delle variabili presenti in un gruppo di variabili originale. È inoltre possibile specificare un nome e un'etichetta per la nuova variabile indice.

Figura 9-28
Ristrutturazione di dati guidata: Creazione di una variabile indice, Passo 5


Ristrutturazione di dati guidata - Passo 5 di 7
Variabili in casi: Crea una variabile indice
Si è scelto di creare una variabile indice. I valori della variabile possono essere numeri sequenziali oppure i nomi delle variabili in un gruppo.
Nella tabella è possibile specificare il nome e l'etichetta per la variabile indice.
Che tipo di valori indice?
Numeri sequenziali
Valori indice: 1, 2, 3
Nomi di variabile
Valori indice: a1,a2,a3
Modifica l'etichetta e il nome della variabile indice:
Nome
Etichetta
Livelli
Valori indice
1
Indice
3
1, 2, 3
< Indietro
Avanti >
Fine
Annulla
?

Per ulteriori informazioni, vedere "Esempio di indice singolo per Variabili in casi" a pag. 244.

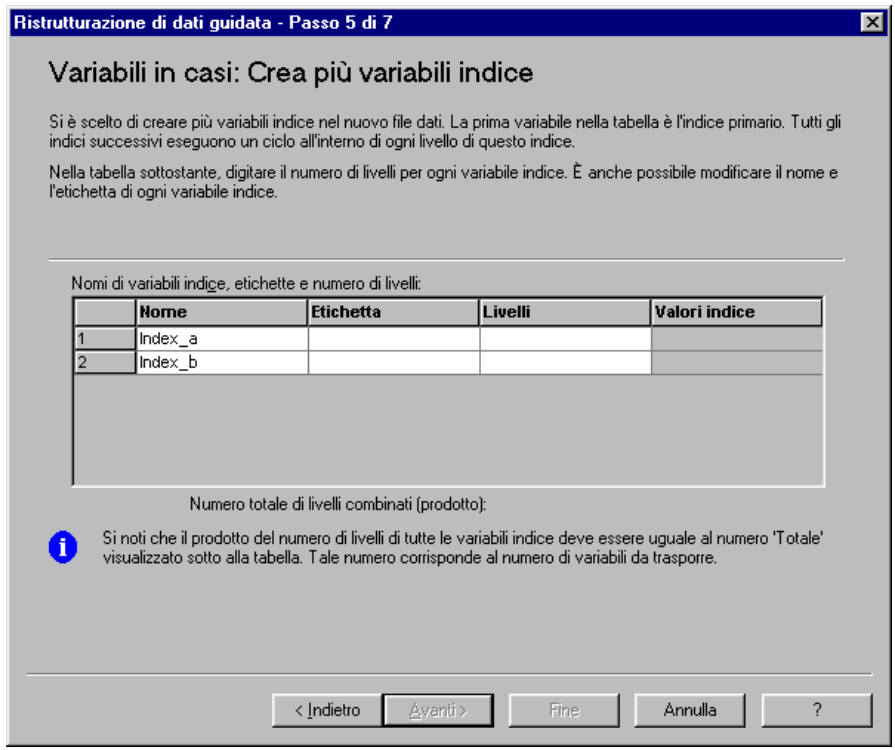
- **Numeri sequenziali.** Come valori dell'indice verranno assegnati automaticamente numeri sequenziali.
- **Nomi delle variabili.** Come valori dell'indice verranno utilizzati i nomi del gruppo di variabili selezionato. Selezionare un gruppo di variabili dall'elenco.
- **Nomi ed etichette.** Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per la variabile indice.

Ristrutturazione di dati guidata (Variabili in casi): Crea più variabili indice

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe e di creare più variabili indice

e consente di specificare il numero di livelli per ogni variabile indice. È inoltre possibile specificare un nome e un'etichetta per la nuova variabile indice.

Figura 9-29
Ristrutturazione di dati guidata: Creazione di più variabili indice, Passo 5


Ristrutturazione di dati guidata - Passo 5 di 7
Variabili in casi: Crea più variabili indice
Si è scelto di creare più variabili indice nel nuovo file dati. La prima variabile nella tabella è l'indice primario. Tutti gli indici successivi eseguono un ciclo all'interno di ogni livello di questo indice.
Nella tabella sottostante, digitare il numero di livelli per ogni variabile indice. È anche possibile modificare il nome e l'etichetta di ogni variabile indice.
Nomi di variabili indice, etichette e numero di livelli:
Nome
Etichetta
Livelli
Valori indice
1
Index_a
2
Index_b
Numero totale di livelli combinati (prodotto):
Si noti che il prodotto del numero di livelli di tutte le variabili indice deve essere uguale al numero 'Totale' visualizzato sotto alla tabella. Tale numero corrisponde al numero di variabili da trasporre.
< Indietro
Avanti >
Fine
Annulla
?

Per ulteriori informazioni, vedere "Esempio di due indici per Variabili in casi" a pag. 245.

Numero di livelli registrati nel file corrente. Prendere in considerazione il numero dei livelli di fattore registrati nei dati correnti. Un **livello** definisce un gruppo di casi per i quali sono valide le stesse condizioni. Se sono presenti più fattori, i dati correnti

devono essere disposti in modo tale che i livelli del primo fattore rappresentino un indice primario all'interno del quale i livelli dei fattori successivi eseguono cicli.

Numero dei livelli nel nuovo file. Specificare il numero dei livelli per ogni indice. I valori per più variabili indice sono sempre rappresentati da numeri sequenziali. I valori iniziano da 1 e vengono incrementati per ogni livello. Il primo indice incrementa il più lento e l'ultimo incrementa il più veloce.

Numero totale di livelli combinati. Non è possibile creare più livelli di quelli presenti nei dati correnti. Poiché i dati ristrutturati conterranno una riga per ogni combinazione di trattamenti, la procedura guidata verifica il numero dei livelli creati e confronta il prodotto dei livelli creati con il numero delle variabili presenti nei gruppi di variabili, che devono corrispondere.

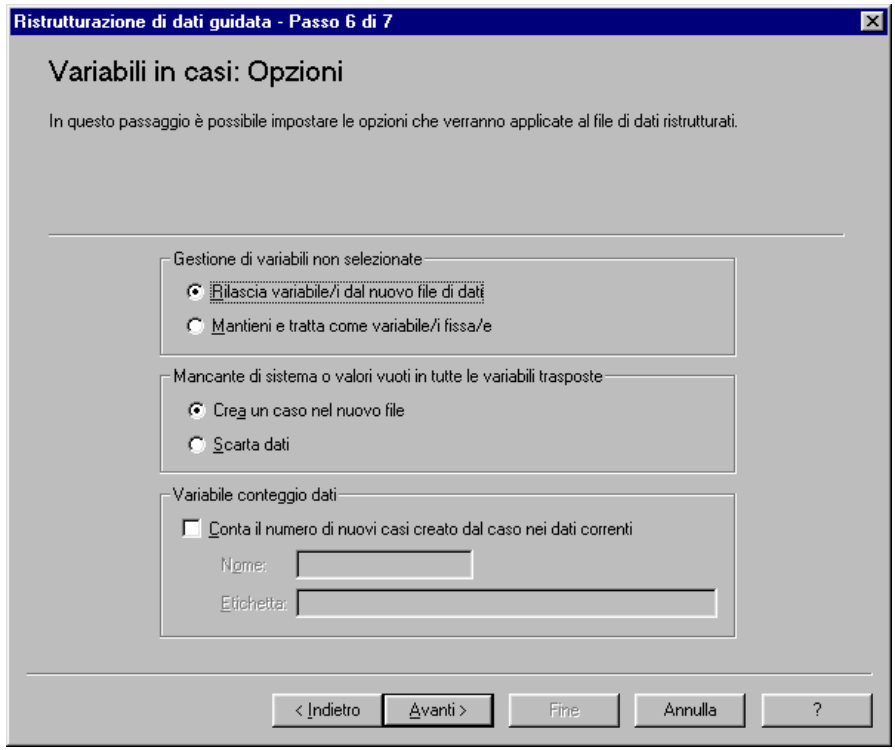
Nomi ed etichette. Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per le variabili indice.

Ristrutturazione di dati guidata (Variabili in casi): Opzioni

Nota: questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di variabili in righe

e consente di specificare le opzioni per il nuovo file ristrutturato.

Figura 9-30
Ristrutturazione di dati guidata: Opzioni, Passo 6


Ristrutturazione di dati guidata - Passo 6 di 7
Variabili in casi: Opzioni
In questo passaggio è possibile impostare le opzioni che verranno applicate al file di dati ristrutturati.
Gestione di variabili non selezionate
Rilascia variabile/i dal nuovo file di dati
Mantieni e tratta come variabile/i fissa/e
Mancante di sistema o valori vuoti in tutte le variabili trasposte
Crea un caso nel nuovo file
Scarta dati
Variabile conteggio dati
Conta il numero di nuovi casi creato dal caso nei dati correnti
Nome:
Etichetta:
< Indietro
Avanti >
Fine
Annulla
?

Esclusione delle variabili non selezionate. Nel passo 3 relativo alla selezione delle variabili sono stati selezionati dai dati correnti i gruppi di variabili da ristrutturare, le variabili da copiare e una variabile di identificazione. I dati delle variabili selezionate verranno visualizzati nel nuovo file. Se nei dati correnti sono presenti altre variabili, è possibile scegliere di escluderle oppure di mantenerle.

Conservazione dei dati mancanti. La procedura guidata verifica se in ogni nuova riga potenziale sono presenti **valori nulli** ovvero valori mancanti di sistema o vuoti. Si può scegliere di mantenere o escludere le righe che contengono valori nulli.

Creazione di una variabile conteggio. È possibile creare nel nuovo file una **variabile conteggio** che contiene il numero delle nuove righe create da una riga nei dati correnti. Le variabili conteggio possono risultare utili se si sceglie di escludere i valori nulli dal nuovo file perché in questo modo è possibile creare un numero diverso

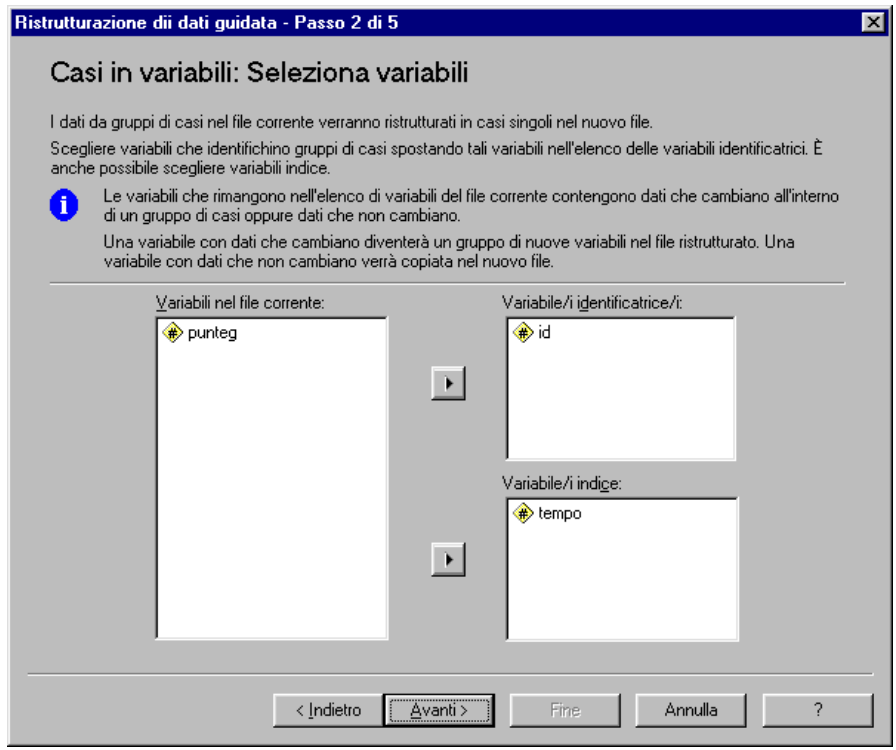
di nuove righe per una riga specifica dei dati correnti. Fare clic su una cella per modificare il nome predefinito della variabile e specificare un'etichetta descrittiva per la variabile conteggio.

Ristrutturazione di dati guidata (Casi in variabili): Selezionare Variabili

Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di casi in colonne

e fornisce informazioni su come utilizzare le variabili del file corrente nel nuovo file.

Figura 9-31
Ristrutturazione di dati guidata: Selezione delle variabili, Passo 2


Ristrutturazione dii dati guidata - Passo 2 di 5
Casi in variabili: Seleziona variabili
I dati da gruppi di casi nel file corrente verranno ristrutturati in casi singoli nel nuovo file.
Scegliere variabili che identifichino gruppi di casi spostando tali variabili nell'elenco delle variabili identificatrici. È anche possibile scegliere variabili indice.
Le variabili che rimangono nell'elenco di variabili del file corrente contengono dati che cambiano all'interno di un gruppo di casi oppure dati che non cambiano.
Una variabile con dati che cambiano diventerà un gruppo di nuove variabili nel file ristrutturato. Una variabile con dati che non cambiano verrà copiata nel nuovo file.
Variabili nel file corrente:
punteg
Variabile/i identificatrice/i:
id
Variabile/i indice:
tempo
< Indietro
Avanti >
Fine
Annulla
?

Identificazione dei gruppi di casi nei dati correnti. Un **gruppo di casi** è un gruppo di righe correlate dal fatto che misurano la stessa unità di osservazione, ad esempio un individuo o un'istituzione. È necessario specificare le variabili del file corrente che identificano i gruppi di casi, in modo tale che sia possibile consolidare ogni gruppo in una singola riga del nuovo file. Spostare le variabili che identificano i gruppi di casi nell'elenco Variabile/i identificatrice/i. Per identificare i gruppi di casi vengono utilizzate le variabili per la suddivisione del file di dati corrente. Ogni volta che viene rilevata una nuova combinazione di valori di identificazione, verrà creata una nuova riga e pertanto i casi del file corrente dovrebbero essere ordinati in base ai valori delle variabili di identificazione nello stesso ordine delle variabili nell'elenco Variabile/i identificatrice/i. Se il file di dati corrente non è già ordinato, è possibile eseguirne l'ordinamento nel passo successivo.

Creazione dei nuovi gruppi di variabili nel nuovo file. Nei dati originali, una variabile viene visualizzata in una singola colonna, mentre nel nuovo file di dati, la stessa variabile viene visualizzata in più colonne nuove. Le **variabili indice** sono le variabili dei dati correnti che vengono utilizzate per creare le nuove colonne. I dati ristrutturati conterranno una nuova variabile per ogni valore univoco di tali colonne. Spostare le variabili che verranno utilizzate per creare i nuovi gruppi di variabili nell'elenco Variabile/i indice. Quando vengono visualizzate le opzioni, è inoltre possibile scegliere di ordinare le nuove colonne in base all'indice.

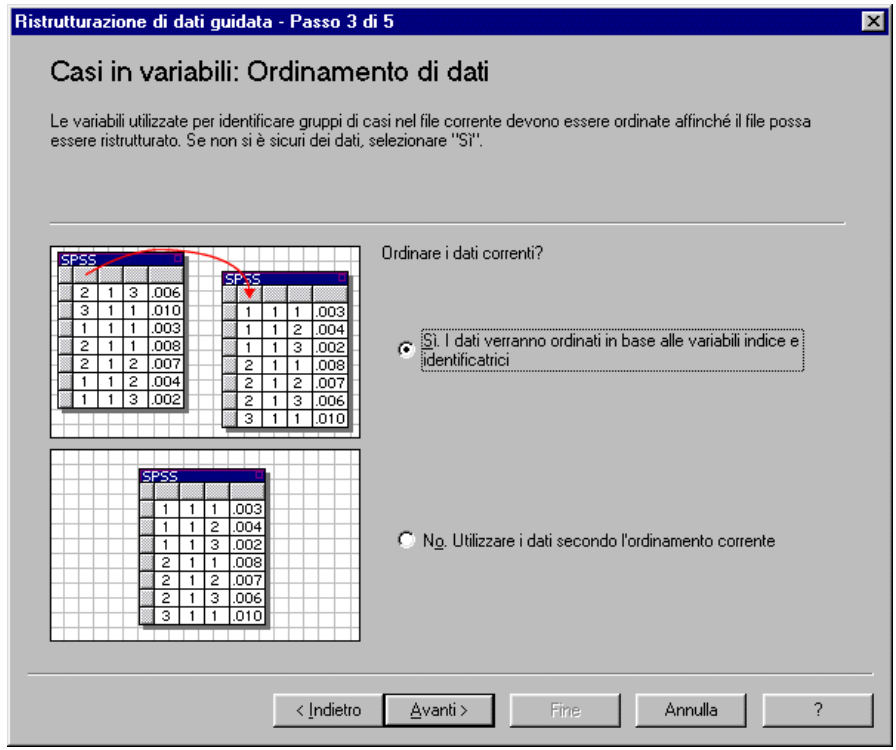
Cosa accade alle altre colonne. La procedura guidata stabilisce automaticamente cosa accade alle altre variabili rimaste nell'elenco relativo al file corrente. Controlla ogni variabile per verificare se i valori dei dati variano all'interno di un gruppo di casi. In caso affermativo, ristruttura i valori in un gruppo di variabili del nuovo file. In caso contrario, copia i valori nel nuovo file.

Ristrutturazione di dati guidata (Casi in variabili): Ordina Dati

Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di casi in colonne

e consente di stabilire se si desidera ordinare il file corrente prima di ristrutturarlo. Ogni volta che viene rilevata una nuova combinazione di valori di identificazione, verrà creata una nuova riga ed è pertanto importante che i dati vengano ordinati in base alle variabili che identificano i gruppi di casi.

Figura 9-32
Ristrutturazione di dati guidata: Ordinamento dei dati, Passo 3


Ristrutturazione di dati guidata - Passo 3 di 5
Casi in variabili: Ordinamento di dati
Le variabili utilizzate per identificare gruppi di casi nel file corrente devono essere ordinate affinché il file possa essere ristrutturato. Se non si è sicuri dei dati, selezionare "Sì".
Ordinare i dati correnti?
Sì. I dati verranno ordinati in base alle variabili indice e identificatrici
No. Utilizzare i dati secondo l'ordinamento corrente
< Indietro
Avanti >
Fine
Annulla
?

Ordinamento delle righe nel file corrente. Prendere in considerazione la modalità di ordinamento dei dati correnti e le variabili utilizzate per identificare i gruppi di casi, specificate nel passo precedente.

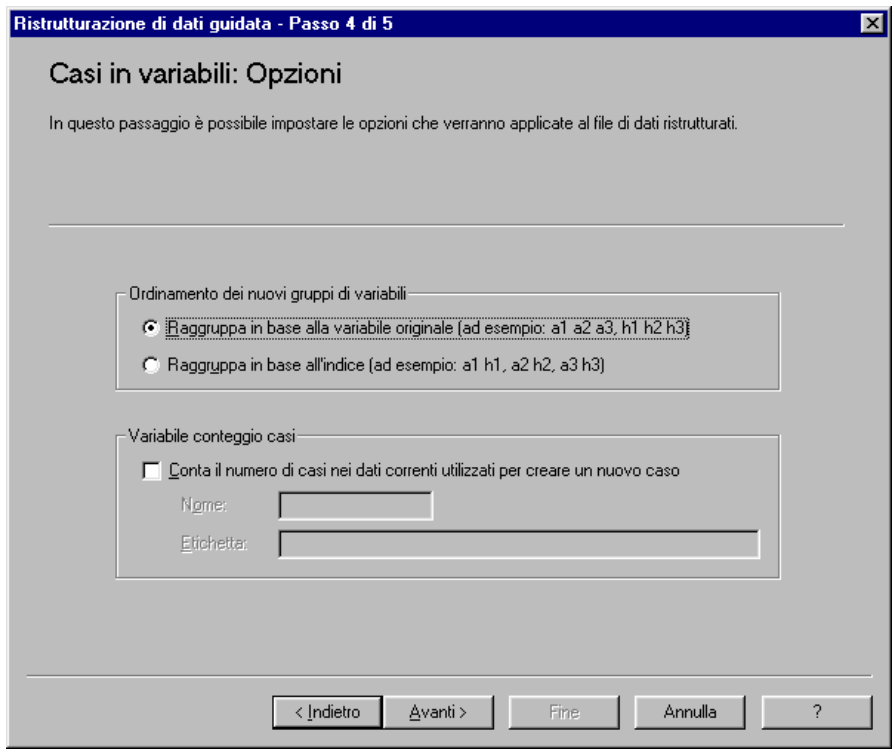
- **Sì.** I dati correnti verranno automaticamente ordinati in base alle variabili di identificazione, nello stesso ordine in cui le variabili vengono visualizzate nell'elenco Variabile/i identificatrice/i del passo precedente. Selezionare questa opzione se i dati non sono ordinati in base alle variabili di identificazione o in caso di incertezze. Questa opzione rende necessario un passaggio distinto dei dati, ma garantisce che le righe vengano ordinate correttamente per la ristrutturazione.
- **No.** I dati correnti non vengono ordinati. Selezionare questa opzione se si è certi che i dati correnti siano ordinati in base alle variabili che identificano i gruppi di casi.

Ristrutturazione di dati guidata (Casi in variabili): Opzioni

Nota: Questo passo viene visualizzato se si sceglie di ristrutturare i gruppi di casi in colonne

e consente di specificare le opzioni per il nuovo file ristrutturato.

Figura 9-33
Ristrutturazione di dati guidata: Opzioni, Passo 4


Ristrutturazione di dati guidata - Passo 4 di 5
Casi in variabili: Opzioni
In questo passaggio è possibile impostare le opzioni che verranno applicate al file di dati ristrutturati.
Ordinamento dei nuovi gruppi di variabili
Raggruppa in base alla variabile originale (ad esempio: a1 a2 a3, h1 h2 h3)
Raggruppa in base all'indice (ad esempio: a1 h1, a2 h2, a3 h3)
Variabile conteggio casi
Conta il numero di casi nei dati correnti utilizzati per creare un nuovo caso
Nome:
Etichetta:
< Indietro
Avanti >
Fine
Annulla
?

Ordinamento dei nuovi gruppi di variabili nel nuovo file

- **Tramite variabile.** Le nuove variabili create da una variabile originale vengono raggruppate insieme.
- **Tramite indice.** Le variabili vengono raggruppate in base ai valori delle variabili indice.

Esempio. Le variabili da ristrutturare sono *a* e *h* e l'indice è *mese*:

a	h	mese

Il raggruppamento tramite variabile restituisce i risultati seguenti:

a.gen	a.feb	h.gen

Il raggruppamento tramite indice restituisce i risultati seguenti:

a.gen	h.gen	a.feb

Creazione di una variabile conteggio. È possibile creare nel nuovo file una variabile conteggio, che contiene il numero delle righe dei dati correnti utilizzate per creare una riga nel nuovo file dati.

Creazione di variabili indicatore. Le variabili indici possono essere utilizzate per creare **variabili indicatore** nel nuovo file di dati. È possibile creare una nuova variabile per ogni valore univoco della variabile indice. Le variabili indicatore indicano la presenza o l'assenza di un valore per un caso. Il valore della variabile indicatore è 1 se il caso contiene un valore, altrimenti è 0.

Esempio. La variabile indice è *prodotto*. Tale variabile consente di registrare i prodotti acquistati da un cliente. I dati originali sono:

cliente	**prodotto**
1	pulcino
1	uova
2	uova
3	pulcino

La creazione di una variabile indicatore risulta in una variabile per ogni valore univoco di *prodotto*. I dati ristrutturati sono:

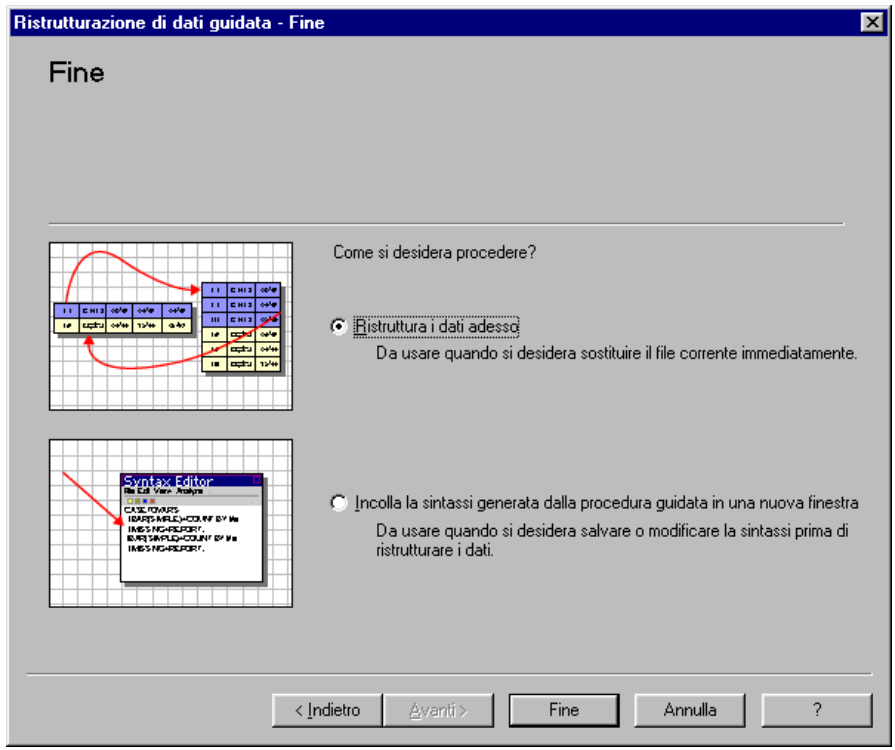
cliente	**indpulcino**	**induova**
1	1	1
2	0	1
3	1	0

In questo esempio, i dati ristrutturati potrebbero essere utilizzati per ottenere i conteggi di frequenze dei prodotti acquistati dai clienti.

Ristrutturazione di dati guidata: Fine

È il passo conclusivo della procedura. Stabilire quale operazione si desidera eseguire per le specifiche.

Figura 9-34
Ristrutturazione di dati guidata: Fine


Ristrutturazione di dati guidata - Fine
Fine
Come si desidera procedere?
Ristruttura i dati adesso
Da usare quando si desidera sostituire il file corrente immediatamente.
Syntax Editor
Incolla la sintassi generata dalla procedura guidata in una nuova finestra
Da usare quando si desidera salvare o modificare la sintassi prima di ristrutturare i dati.
< Indietro
Avanti >
Fine
Annulla
?

- **Ristruttura ora.** Verrà creato il nuovo file ristrutturato. Specificare se il file corrente deve essere sostituito immediatamente. *Nota*: Se i dati originali sono ponderati, lo saranno anche i nuovi dati, a meno che la variabile utilizzata come peso non sia ristrutturata o esclusa dal nuovo file.
- **Incolla sintassi.** La sintassi generata viene incollata in una finestra di sintassi. Selezionare questa opzione se non si è certi di voler sostituire il file corrente oppure se si desidera modificare la sintassi o salvarla per modificarla in seguito.

Capitolo
10

Utilizzo dell'output

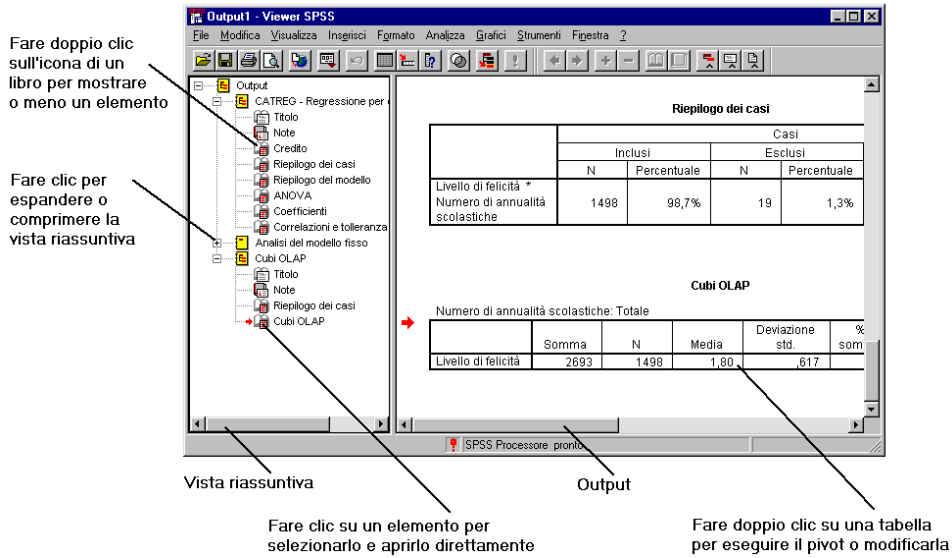
Quando si esegue una procedura, i risultati vengono visualizzati in una finestra denominata Viewer. In questa finestra è possibile spostarsi rapidamente verso qualsiasi parte dell'output che si desidera visualizzare. È inoltre possibile gestire l'output e creare un documento che contenga esattamente l'output desiderato.

Viewer

I risultati vengono visualizzati nel Viewer. È possibile utilizzare il Viewer per:

- Spostarsi nei risultati.
- Visualizzare o nascondere le tabelle e i grafici selezionati.
- Modificare l'ordine di visualizzazione dell'output spostando gli elementi selezionati.
- Spostare elementi tra il Viewer e altre applicazioni.

Figura 10-1
Viewer



Il Viewer è suddiviso in due riquadri:

- Il riquadro sinistro contiene una vista riassuntiva del contenuto dell'output.
- Il riquadro destro contiene tabelle statistiche, grafici e output testuale.

È possibile utilizzare le barre di scorrimento per visualizzare i risultati oppure fare clic su un elemento nella vista riassuntiva per passare direttamente alla tabella o al grafico corrispondente. È possibile fare clic e trascinare il bordo destro della vista riassuntiva per modificarne la larghezza.

Utilizzo della Bozza

Se si desidera un output testuale semplice piuttosto che le tabelle pivot interattive, è possibile utilizzare la Bozza.

Per utilizzare la Bozza:

▶ Dai menu di qualsiasi finestra, scegliere:

Modifica
  Opzioni...

- Nella scheda Generale, fare clic su Bozza per il tipo di output.
- Per modificare le opzioni relative al formato dell'output della Bozza, fare clic sulla scheda Bozza.

  Per ulteriori informazioni, vedere "Bozza" in Capitolo 11 a pag. 291. Per ulteriori informazioni, è possibile anche consultare la Guida:

- Dai menu di qualsiasi finestra, scegliere:

  Aiuto
  Argomenti

- Fare clic sulla scheda Indice nella finestra dell'Aiuto in linea.
- Digitare bozza e fare doppio clic sulla voce dell'indice.

Attivazione e disattivazione della visualizzazione dell'output

Nel Viewer è possibile visualizzare o nascondere in modo selettivo singole tabelle o risultati inclusi in una procedura. Questo processo risulta utile se si desidera visualizzare solo una parte dell'output.

Per nascondere tabelle e grafici

- Fare doppio clic sull'icona a forma di libro dell'elemento corrispondente nella vista riassuntiva del Viewer.

  oppure

- Fare clic sull'elemento per selezionarlo.
- Dai menu, scegliere:

  Visualizza
  Nascondi

  oppure

- Fare clic sull'icona a forma di libro chiuso (Nascondi) sulla barra degli strumenti.

  Verrà attivata l'icona a forma di libro aperto (Visualizza), ad indicare che l'elemento è nascosto.

Per nascondere i risultati della procedura

- Fare clic sulla casella a sinistra del nome della procedura nella vista riassuntiva.

  Verranno nascosti tutti i risultati della procedura e la vista riassuntiva risulterà compressa.

Spostamento, copia ed eliminazione dell'output

È possibile riorganizzare l'output copiando, spostando o eliminando un elemento o un gruppo di elementi.

Per spostare l'output nel Viewer

- Fare clic su un elemento nella vista riassuntiva. Fare clic tenendo premuto Maiusc per selezionare più elementi oppure fare clic tenendo premuto Ctrl per selezionare elementi non contigui.
- Trascinare gli elementi selezionati (tenendo premuto il pulsante del mouse durante l'operazione di trascinamento).
- Rilasciare il pulsante del mouse nella posizione in cui si desidera inserire gli elementi spostati.

  È inoltre possibile spostare gli elementi utilizzando i comandi Taglia e Incolla dopo del menu Modifica.

Per eliminare l'output dal Viewer

- Fare clic su un elemento nella vista riassuntiva. Fare clic tenendo premuto Maiusc per selezionare più elementi oppure fare clic tenendo premuto Ctrl per selezionare elementi non contigui.
- Premere Canc.

  oppure

- Dai menu, scegliere:

  Modifica
  Elimina

Per copiare output nel Viewer

- Fare clic su un elemento nella vista riassuntiva. Fare clic tenendo premuto Maiusc per selezionare più elementi oppure fare clic tenendo premuto Ctrl per selezionare elementi non contigui.

- Tenere premuto Ctrl mentre si utilizza il mouse per trascinare gli elementi selezionati. Tenere premuto il pulsante del mouse durante il trascinamento.

- Rilasciare il pulsante del mouse per inserire gli elementi nella posizione desiderata.

  È inoltre possibile copiare gli elementi utilizzando i comandi Copia e Incolla dopo del menu Modifica.

Modifica dell'allineamento iniziale

Per impostazione predefinita, inizialmente tutto l'output è allineato a sinistra. È possibile modificare l'allineamento iniziale (comando Opzioni del menu Modifica, scheda Viewer) in qualsiasi momento.

Modifica dell'allineamento degli elementi dell'output

- Selezionare la vista riassuntiva, quindi fare clic sugli elementi da allineare Fare clic tenendo premuto Maiusc o Ctrl per selezionare più elementi.

- Dai menu, scegliere:
  Formato
  Allinea a sinistra

  Altre opzioni di allineamento sono Centra e Allinea a destra.

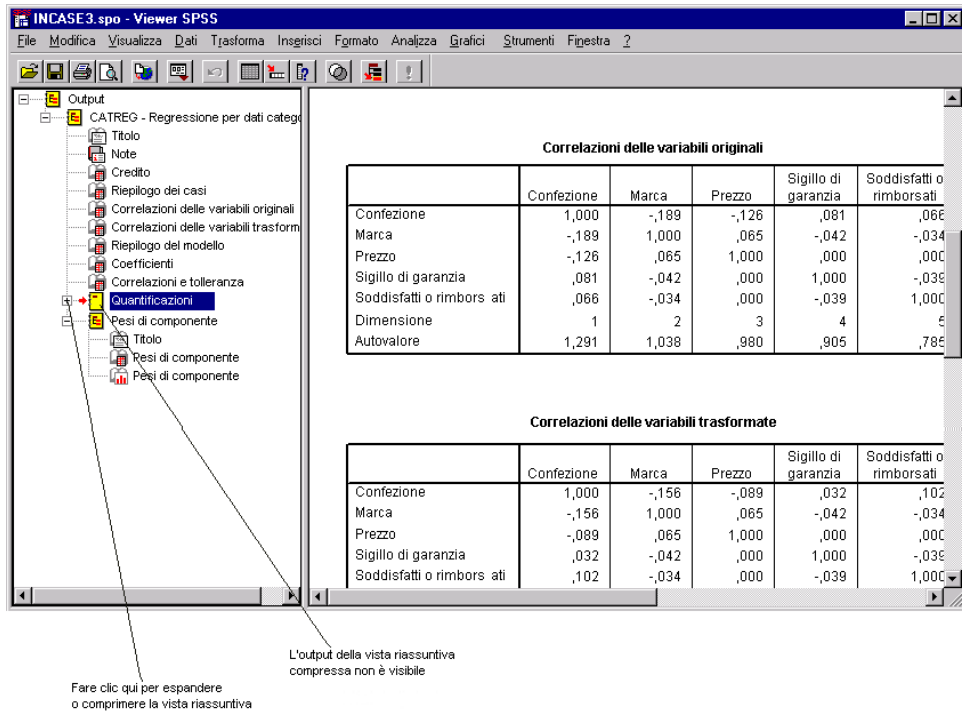
  Nota: Nel Viewer tutti i risultati vengono allineati a sinistra. Le opzioni di allineamento influiscono solo sull'allineamento dell'output stampato. Gli elementi centrati e allineati a destra sono contrassegnati da un piccolo simbolo situato sopra e a sinistra di ciascun elemento.

Vista riassuntiva del Viewer

La vista riassuntiva include un sommario del documento del Viewer. È possibile utilizzare la vista riassuntiva per spostarsi nell'output ed impostarne la visualizzazione. La maggior parte delle operazioni eseguite nella vista riassuntiva vengono applicate all'output.

- Gli elementi selezionati nella vista riassuntiva verranno visualizzati anche nell'output.
- Gli elementi spostati nella vista riassuntiva verranno spostati anche nell'output.
- Se la vista riassuntiva viene compressa, l'output di tutti gli elementi ai livelli compressi verrà nascosto.

Figura 10-2
Vista riassuntiva compressa e output nascosto


INCASE3.spo - Viewer SPSS
File Modifica Visualizza Dati Trasforma Inserisci Formato Analizza Grafici Strumenti Finestra ?
Output
CATREG - Regressione per dati categ
Titolo
Note
Credito
Riepilogo dei casi
Correlazioni delle variabili originali
Correlazioni delle variabili trasform
Riepilogo del modello
Coefficienti
Correlazioni e tolleranza
Quantificazioni
Pesi di componente
Titolo
Pesi di componente
Pesi di componente
Correlazioni delle variabili originali
Confezione Marca Prezzo Sigillo di garanzia Soddisfatti o rimborsati
Confezione 1,000 -,189 -,126 ,081 ,066
Marca -,189 1,000 ,065 -,042 -,034
Prezzo -,126 ,065 1,000 ,000 ,000
Sigillo di garanzia ,081 -,042 ,000 1,000 -,039
Soddisfatti o rimbors ati ,066 -,034 ,000 -,039 1,000
Dimensione 1 2 3 4
Autovalore 1,291 1,038 ,980 ,905
Correlazioni delle variabili trasformate
Confezione Marca Prezzo Sigillo di garanzia Soddisfatti o rimborsati
Confezione 1,000 -,156 -,089 ,032 ,102
Marca -,156 1,000 ,065 -,042 -,034
Prezzo -,089 ,065 1,000 ,000 ,000
Sigillo di garanzia ,032 -,042 ,000 1,000 -,039
Soddisfatti o rimbors ati ,102 -,034 ,000 -,039 1,000
L'output della vista riassuntiva compressa non è visibile
Fare clic qui per espandere o comprimere la vista riassuntiva

Impostazione della visualizzazione dell'output. Per impostare le modalità di visualizzazione dell'output è possibile:

- Espandere e comprimere la vista riassuntiva.
- Modificare il livello degli elementi selezionati nella vista riassuntiva.
- Modificare le dimensioni degli elementi nella vista riassuntiva.
- Modificare il tipo di carattere utilizzato nella vista riassuntiva.

Per comprimere ed espandere la vista riassuntiva

▶ Fare clic sulla casella a sinistra dell'elemento che si desidera comprimere o espandere nella vista riassuntiva.

oppure

▶ Fare clic sull'elemento nella vista riassuntiva.

▶ Dai menu, scegliere:

Visualizza
  Comprimi

oppure

Visualizza
  Espandi

Per modificare il livello degli elementi

▶ Fare clic sull'elemento nella vista riassuntiva.

▶ Fare clic sulla freccia rivolta verso sinistra sulla barra degli strumenti della vista riassuntiva per spostare l'elemento avanti di un livello (a sinistra).

▶ Fare clic sulla freccia rivolta verso destra sulla barra degli strumenti della vista riassuntiva per spostare l'elemento indietro di un livello (a destra).

oppure

Dai menu, scegliere:

Modifica
  Output
    Promuovi

oppure

Modifica
  Output
    Retrocedi

La modifica del livello risulta particolarmente utile dopo aver spostato gli elementi nella vista riassuntiva. Spostando alcuni elementi è possibile modificare il livello degli elementi nella vista riassuntiva. Per ripristinare i livelli originali, utilizzare i pulsanti freccia destra e sinistra sulla barra degli strumenti.

Per modificare le dimensioni degli elementi nella vista riassuntiva

▶ Dai menu, scegliere:

Visualizza
  Dimensioni vista riassuntiva
    Piccola

Altre opzioni sono Media e Grande.

Le icone e il testo corrispondente cambieranno dimensioni.

Per modificare il carattere nella vista riassuntiva

▶ Dai menu, scegliere:

Visualizza
  Carattere vista riassuntiva...

▶ Selezionare un carattere.

Aggiunta di elementi nel Viewer

Nel Viewer è possibile inserire elementi quali titoli, nuovo testo, grafici o materiali di altre applicazioni.

Per aggiungere un nuovo titolo o elemento testuale al contenuto del Viewer

Nel Viewer è possibile inserire elementi testuali che non siano collegati a una tabella o a un grafico.

▶ Fare clic sulla tabella, sul grafico o sull'oggetto che precederà il titolo o il testo.

- Dai menu, scegliere:

  Inserisci
  Nuovo titolo

  oppure

  Inserisci
  Nuovo testo

- Fare doppio clic sul nuovo oggetto.
- Inserire il testo desiderato in questa posizione.

Per aggiungere un file di testo

- Nell'output o nella vista riassuntiva del Viewer, fare clic sulla tabella, sul grafico o sull'oggetto che precederà il testo.
- Dai menu, scegliere:

  Inserisci
  File di testo...

- Selezionare un file di testo.

  Fare doppio clic sul testo da modificare.

Utilizzo dell'output in altre applicazioni

Le tabelle pivot e i grafici possono essere copiati e incollati in altre applicazioni Windows, ad esempio in elaboratori di testo o fogli elettronici. È possibile incollare le tabelle pivot o i grafici in diversi formati, tra cui:

Oggetto incorporato. È possibile incorporare le tabelle pivot e i grafici interattivi nelle applicazioni che supportano gli oggetti ActiveX. Dopo aver incollato la tabella è possibile attivarla nell'applicazione di destinazione facendo doppio clic su di essa e quindi modificarla in base alle procedure del Viewer.

Illustrazioni (metafile). È possibile incollare le tabelle pivot, l'output testuale e i grafici come illustrazioni metafile. È possibile ridimensionare il formato dell'immagine nell'altra applicazione. A volte è possibile eseguire alcune modifiche utilizzando gli strumenti disponibili nell'altra applicazione. Le tabelle pivot incollate come illustrazioni mantengono le caratteristiche dei bordi e del carattere.

Formato RTF (Rich Text Format). È possibile incollare le tabelle pivot in formato RTF in altre applicazioni. Nella maggior parte delle applicazioni, la tabella pivot viene incollata come una normale tabella e può essere quindi modificata.

Bitmap. È possibile incollare i grafici in altre applicazioni come bitmap.

BIFF. È possibile incollare il contenuto di una tabella in un foglio elettronico mantenendo la precisione numerica.

Testo. È possibile copiare il contenuto di una tabella e incollarlo come testo. Questo processo può essere utile in applicazioni come la posta elettronica che possono ricevere o trasmettere solo testo.

Per copiare una tabella o un grafico

- Selezionare la tabella o il grafico da copiare.
- Dai menu, scegliere:

Modifica
  Copia

Per copiare e incollare l'output in un'altra applicazione

- Copiare l'output nel Viewer.
- Dai menu dell'applicazione di destinazione, scegliere:

Modifica
  Incolla

oppure

Modifica
  Incolla speciale...

Incolla. L'output viene copiato negli Appunti in diversi formati. Ciascuna applicazione determina il formato migliore da utilizzare con il comando Incolla. In numerose applicazioni utilizzando il comando Incolla l'output verrà incollato come immagine (metafile). Negli elaboratori di testo con il comando Incolla è possibile incollare le tabelle pivot in formato RTF, ovvero come tabelle. Nei fogli elettronici il comando Incolla consente di incollare le tabelle pivot in formato BIFF. I grafici vengono incollati come metafile.

Incolla speciale. L'output viene copiato negli Appunti in diversi formati. Il comando Incolla speciale consente di selezionare il formato desiderato dall'elenco dei formati disponibili nell'applicazione di destinazione.

Per incorporare una tabella in un'altra applicazione

È possibile incorporare le tabelle pivot e i grafici interattivi in altre applicazioni utilizzando il formato ActiveX. Un oggetto incorporato può essere attivato nell'applicazione di destinazione facendo doppio clic su di esso ed è quindi possibile modificarlo e applicarvi il Embedding come nel Viewer.

Se si dispone di applicazioni che supportano gli oggetti ActiveX:

- Fare doppio clic su *objs-on.bat*, situato nella directory di installazione di SPSS.

  In questo modo verrà attivata la funzione di incorporamento di oggetti ActiveX per le tabelle pivot. Il file *objs-off.bat* consente di disattivare l'incorporamento di oggetti ActiveX.

Per incorporare una tabella pivot o un grafico interattivo in un'altra applicazione:

- Nel Viewer copiare la tabella.
- Dai menu dell'applicazione di destinazione, scegliere:

  Modifica
  Incolla speciale...

- Dall'elenco selezionare Oggetto tabella pivot SPSS *o* Oggetto controllo grafici SPSS.

  L'applicazione di destinazione deve supportare gli oggetti ActiveX. Per informazioni sul supporto degli oggetti ActiveX, vedere la documentazione dell'applicazione. Alcune applicazioni che non supportano gli oggetti ActiveX possono accettare inizialmente le tabelle pivot ActiveX, ma è comunque possibile che in seguito producano effetti indesiderati. È consigliabile utilizzare le tabelle pivot incorporate solo dopo aver verificato la stabilità dell'applicazione con gli oggetti ActiveX incorporati.

Per incollare una tabella o un grafico come immagine

- Nel Viewer copiare la tabella o il grafico.

- Dai menu dell'applicazione di destinazione, scegliere:
  Modifica
  Incolla speciale...

- Scegliere Immagine dall'elenco.

  L'elemento verrà incollato come metafile. Nel metafile saranno disponibili solo gli strati e le colonne che risultavano visibili quando l'elemento è stato copiato. Gli altri strati o le colonne nascoste non saranno disponibili.

Per incollare una tabella pivot come tabella

- Nel Viewer copiare la tabella.

- Dai menu dell'applicazione di destinazione, scegliere:
  Modifica
  Incolla speciale...

- Dall'elenco selezionare l'opzione relativa al testo in formato RTF (Rich Text Format).

  La tabella pivot verrà incollata come una tabella. Nella tabella verranno incollati solo lo strato e le colonne che risultavano visibili quando l'elemento è stato copiato. Gli altri strati o le colonne nascoste non saranno disponibili. In questo formato, è possibile copiare e incollare solo una tabella pivot alla volta.

Per incollare una tabella pivot come testo

- Nel Viewer copiare la tabella.

- Dai menu dell'applicazione di destinazione, scegliere:
  Modifica
  Incolla speciale...

- Dall'elenco selezionare Testo non formattato.

  Le tabelle pivot non formattate contengono tabulazioni tra le colonne. È possibile allineare le colonne modificando le tabulazioni nell'altra applicazione.

Per copiare e incollare più elementi in un'altra applicazione

- Selezionare le tabelle e/o i grafici da copiare. Per selezionare più elementi fare clic tenendo premuto Maiusc o Ctrl.
- Dai menu, scegliere:

  Modifica
  Copia oggetti
- Dai menu dell'applicazione di destinazione, scegliere:

  Modifica
  Incolla

  Nota: utilizzare il comando Copia oggetti solo per copiare più elementi dal Viewer in un'altra applicazione. Per eseguire operazioni di copia e incolla all'interno di documenti del Viewer, ad esempio tra due finestre del Viewer, scegliere Copia dal menu Modifica.

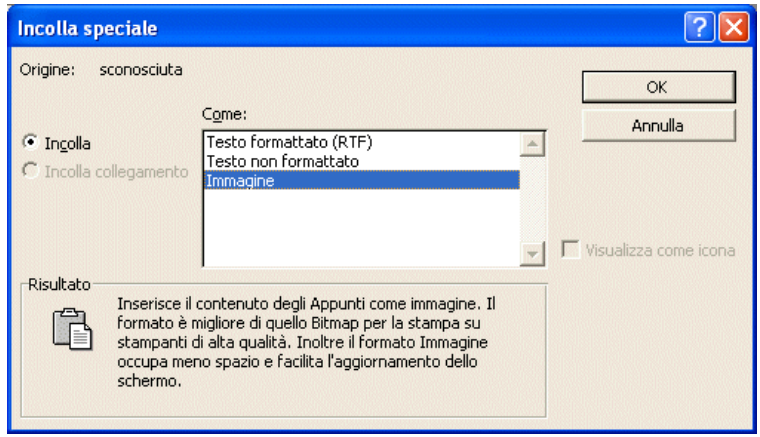
Per incollare oggetti nel Viewer

Nel Viewer possono essere incollati oggetti di altre applicazioni. È possibile utilizzare i comandi Incolla dopo o Incolla speciale. Entrambi i comandi consentono di incollare il nuovo oggetto dopo l'oggetto selezionato nel Viewer. Utilizzare il comando Incolla speciale per scegliere il formato dell'oggetto incollato.

Incolla speciale

Il comando Incolla speciale consente di selezionare il formato di un oggetto copiato che viene incollato nel Viewer. I tipi di file consentiti per gli oggetti sono elencati nella finestra di dialogo Incolla speciale. L'oggetto verrà inserito nel Viewer dopo l'oggetto selezionato.

Figura 10-3
Finestra di dialogo Incolla speciale


Incolla speciale
Origine: sconosciuta
OK
Annulla
Come:
Incolla
Incolla collegamento
Testo formattato (RTF)
Testo non formattato
Immagine
Visualizza come icona
Risultato
Inserisce il contenuto degli Appunti come immagine. Il formato è migliore di quello Bitmap per la stampa su stampanti di alta qualità. Inoltre il formato Immagine occupa meno spazio e facilita l'aggiornamento dello schermo.

Per incollare oggetti da altre applicazioni nel Viewer

- Nelle altre applicazioni, copiare semplicemente l'oggetto.
- Nell'output o nella vista riassuntiva del Viewer, fare clic sulla tabella, sul grafico o sull'oggetto che precederà l'oggetto.
- Dai menu, scegliere:

  Modifica
  Incolla speciale...

- Selezionare il formato dell'oggetto desiderato dall'elenco.

Esporta output

L'opzione Esporta output consente di salvare tabelle pivot e output testuale in formato HTML, testo, Word/RTF, Excel e PowerPoint (è richiesto PowerPoint 97 o versioni successive) e di salvare i grafici in diversi formati comuni utilizzati da altre applicazioni. *Nota*: l'opzione Esporta in PowerPoint non è disponibile nella versione per studenti.

Output. Consente di esportare qualsiasi combinazione di tabelle pivot, output testuale e grafici.

- Per i formati HTML e di testo, i grafici vengono esportati nel formato di esportazione correntemente selezionato. Per il formato HTML, i grafici vengono incorporati come indirizzi e possono essere esportati in un formato adatto all'inserimento in documenti HTML. Per il formato di testo, nel file di testo viene inserita una linea per ogni grafico in cui viene indicato il nome di file del grafico esportato.
- Per il formato Word/RTF, i grafici vengono esportati in formato metafile di Windows e incorporati nel documento Word.
- I grafici non vengono inclusi in documenti Excel.
- Nel formato PowerPoint, i grafici vengono esportati in formato TIFF ed incorporati nel file PowerPoint.

Output (senza grafici). Consente di esportare tabelle pivot e output testuale. I grafici nel Viewer vengono ignorati.

Solo grafici. Sono disponibili diversi formati di esportazione: metafile di Windows (WMF), metafile avanzati (EMF),bitmap di Windows (BMP), PostScript encapsulated (EPS), JPEG, TIFF, PNG e PICT di Macintosh.

Esporta. È possibile esportare tutti gli oggetti nel Viewer, tutti gli oggetti visibili oppure solo gli oggetti selezionati.

Esporta formato. Per l'output, le opzioni disponibili sono HTML, testo, Word/RTF, Excel e PowerPoint. Per il formato testo e HTML, i grafici vengono esportati nel formato dei grafici correntemente selezionato nella finestra di dialogo Opzioni relativa al formato selezionato. Per l'opzione Solo grafici, selezionare un formato di esportazione dall'elenco a discesa. Per i documenti di output, le tabelle e il testo vengono esportati nel seguente modo:

- **File HTML (*.htm).** Le tabelle pivot vengono esportate come tabelle HTML. L'output testuale è esportato come HTML preformattato.
- **File di testo (*.txt).** È possibile esportare le tabelle pivot in formato delimitato da tabulazioni o da spazi. Tutto l'output testuale viene esportato in formato delimitato da spazi.
- **File Excel (*.xls).** Le righe, le colonne e le celle delle tabelle pivot vengono esportate come righe, colonne e celle di Excel, con tutti gli attributi di formattazione, ad esempio, con i bordi delle celle, gli stili del carattere, il colore di sfondo delle celle e così via. L'output testuale viene esportato con tutti gli

attributi del carattere. Ogni riga nell'output testuale è una riga del file di Excel, con tutti i contenuti di una riga raccolti in una singola cella.

- **File Word/RTF (*.doc).** Le tabelle pivot vengono esportate come tabelle Word con tutti gli attributi di formattazione; ad esempio bordi delle celle, stili del carattere, colori dello sfondo e così via. L'output testuale viene esportato come HTML formattato. L'output testuale in SPSS viene sempre visualizzato con caratteri a spaziatura fissa (uno spazio) e viene esportato con gli stessi attributi del carattere. Per il corretto allineamento dell'output testuale delimitato da spazi è necessario utilizzare un carattere a spaziatura fissa (proporzionale).
- **File PowerPoint (*.ppt).** Le tabelle pivot vengono esportate come tabelle Word ed incorporate in diapositive diverse del file PowerPoint; una diapositiva per ciascuna tabella pivot. Tutti gli attributi di formattazione della tabella pivot - quali i bordi delle celle, gli stili dei caratteri, i colori di sfondo, ecc. - vengono mantenuti. L'output testuale viene esportato come HTML formattato. L'output testuale in SPSS viene sempre visualizzato con caratteri a spaziatura fissa (uno spazio) e viene esportato con gli stessi attributi del carattere. Per il corretto allineamento dell'output testuale delimitato da spazi è necessario utilizzare un carattere a spaziatura fissa (proporzionale). *Nota*: l'opzione Esporta in PowerPoint non è disponibile nella versione per studenti.

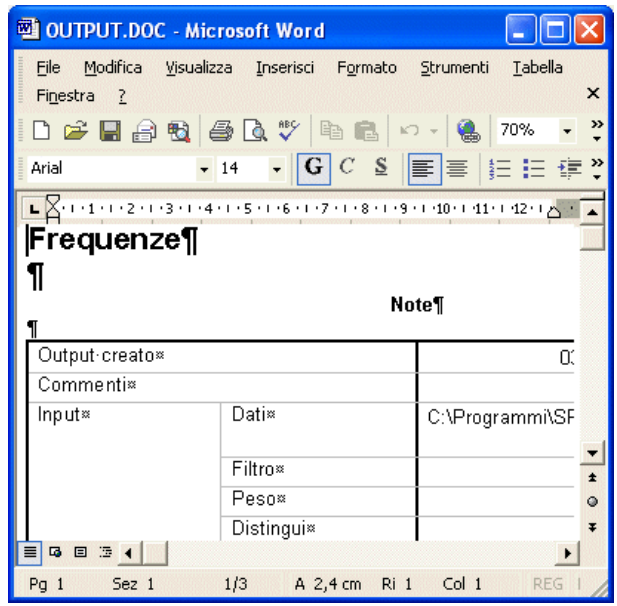
Sistema di gestione dell'output. È possibile esportare automaticamente tutto l 'output oppure solo i tipi di output specificati dall'utente in file di dati in formato di testo, HTML, XML e SPSS. Per ulteriori informazioni, vedere "Sistema di gestione dell'output" in Capitolo 48 a pag. 711.

Per esportare l'output

- ▶ Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- ▶ Dai menu, scegliere:
  File
  Esporta...
- ▶ Specificare un nome di file (o il prefisso di un grafico) e selezionare un formato di esportazione.

Figura 10-4
Finestra di dialogo Esporta output

Figura 10-5
Output esportato in formato Word/RTF



Opzioni HTML, Word/RTF ed Excel

Quando si seleziona Opzioni nella finestra di dialogo Esporta output, viene visualizzata una finestra di dialogo che permette di decidere se includere i titoli e le note a piè di pagina per i documenti esportati in HTML, Word/RTF ed Excel, di specificare le opzioni di esportazione dei grafici per i documenti HTML e di gestire le tabelle pivot con più strati.

Formato immagine. Consente di controllare il formato di esportazione dei grafici e altre impostazioni facoltative, ad esempio la dimensione dei grafici per i documenti HTML. Per il formato Word/RTF, tutti i grafici vengono esportati nel formato metafile di Windows (WMF). In Excel i grafici non vengono inclusi.

Esporta didascalie e titoli. Selezionare questa casella di controllo per includere le note a piè di pagina e le didascalie nell'esportazione delle tabelle pivot.

Esporta tutti gli strati. Selezionare questa casella per esportare tutti gli strati di una tabella pivot. Se la casella di controllo non viene selezionata, viene esportato solo lo strato superiore.

Per impostare le opzioni di esportazione HTML, Word/RTF ed Excel

- Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- Dai menu, scegliere:
  File
  Esporta...
- Selezionare File HTML, File Word/RTF o File Excel come formato di esportazione.
- Fare clic su Opzioni.

Opzioni PowerPoint

Opzioni PowerPoint permette di decidere se includere o meno titoli nelle diapositive, includere piè di pagina e didascalie delle tabelle pivot, come gestire le tabelle pivot multistrato e di specificare le opzioni di esportazione dei grafici per l'esportazione in formato PowerPoint. *Nota*: l'opzione Esporta in PowerPoint non è disponibile nella versione per studenti.

Figura 10-6
Finestra di dialogo Opzioni PowerPoint

Includi titolo nella diapositiva. Selezionare questa casella di controllo per includere un titolo in ciascuna diapositiva creata con l'esportazione. Ciascuna diapositiva contiene un'unica voce esportata da Viewer. Il titolo è costituito dalla voce corrispondente all'elemento desiderato nella vista riassuntiva di Viewer.

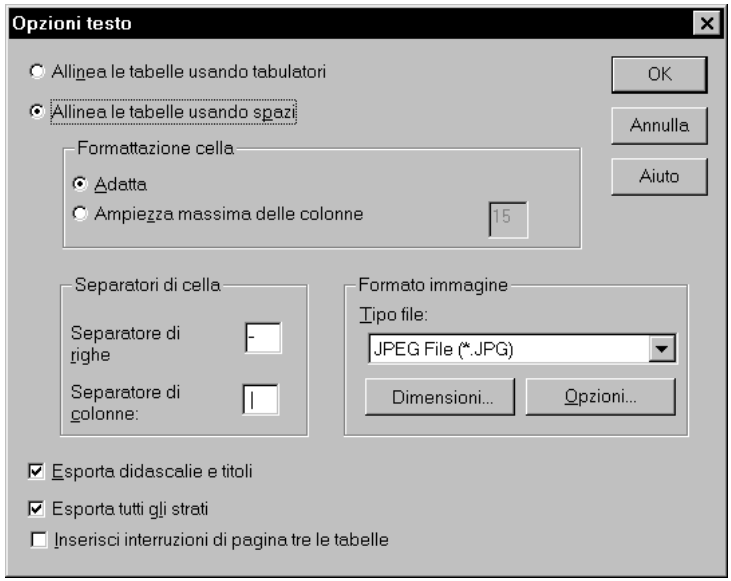
Esporta didascalie e titoli. Selezionare questa casella di controllo per includere le note a piè di pagina e le didascalie nell'esportazione delle tabelle pivot.

Esporta tutti gli strati. Selezionare questa casella per esportare tutti gli strati di una tabella pivot multistrato; ciascuno strato viene inserito in una diapositiva diversa. A tutti gli strati viene assegnato lo stesso titolo. Se la casella di controllo non viene selezionata, viene esportato solo lo strato superiore.

Opzioni testo

Nella finestra di dialogo Opzioni testo è possibile impostare le opzioni per le tabelle pivot, l'output testuale e i formati dei grafici e per l'inserimento di note a piè di pagina e didascalie nei documenti esportati in formato di testo.

Figura 10-7
Finestra di dialogo Opzioni testo



È possibile esportare le tabelle pivot in formato delimitato da tabulazioni o da spazi. Con il formato delimitato da tabulazioni, se una cella non è vuota, verranno stampati il contenuto e un carattere di tabulazione. Se una cella è vuota verrà stampato solo il carattere di tabulazione.

Tutto l'output testuale viene esportato in formato delimitato da spazi. Per il corretto allineamento dell'output testuale delimitato da spazi è necessario utilizzare un carattere a spaziatura fissa (proporzionale).

Formattazione cella. Per le tabelle pivot delimitate da spazi, per default viene eliminato il ritorno a capo automatico e per ciascuna colonna viene impostata la larghezza dell'etichetta o del valore più lungo in essa contenuto. Per limitare la larghezza delle colonne e impostare il ritorno a capo automatico per le etichette più lunghe, specificare il numero di caratteri corrispondente alla larghezza desiderata. Questa impostazione viene applicata solo all'output delle tabelle pivot.

Separatori di cella. Per le tabelle pivot delimitate da spazi è possibile specificare i caratteri utilizzati per creare i bordi delle celle.

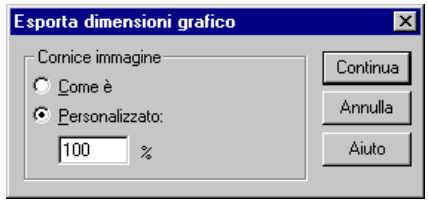
Formato immagine. Consente di impostare il formato di esportazione dei grafici e altre opzioni facoltative quali la dimensione dei grafici.

Inserisci interruzione di pagina tra le tabelle. Inserisce un'interruzione di pagina tra ciascuna tabella. Nel caso delle tabelle pivot multi-strato, inserisce un'interruzione di pagina tra ciascuno strato.

Opzioni per le dimensioni dei grafici

L'opzione Dimensioni consente di controllare la dimensione dei grafici esportati. La specifica della percentuale personalizzata consente di diminuire o di aumentare le dimensioni del grafico esportato fino al 200%.

Figura 10-8
Finestra di dialogo Esporta dimensioni grafico



Per impostare le dimensioni dei grafici esportati

- Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- Dai menu, scegliere:
  File
  Esporta...
- Per l'output, fare clic su Opzioni, selezionare il formato di esportazione e fare clic su Dimensioni.
- Per l'opzione Solo grafici, selezionare il formato di esportazione e fare clic su Dimensioni.

Opzioni di esportazione dei grafici JPEG

Profondità colore. È possibile esportare i grafici JPEG con il formato 16,8 milioni di colori (24 bit) o 256 gradazioni di grigio.

Spazio colore. Questa opzione fa riferimento al modo in cui i colori vengono codificati nell'immagine. Il modello Colore YUV è un tipo di codifica del colore normalmente utilizzato per la trasmissione di file mpeg e video digitale. L'acronimo sta per segnale Y, segnale U, segnale V. Il componente Y specifica la luminanza o gradazione di grigi e i componenti U e V corrispondono alla crominanza (informazioni sul colore).

I rapporti rappresentano i tassi di campionamento per ciascun componente. Riducendo i tassi di campionamento U e V si riduce la dimensione del file (ma anche la qualità). L'opzione spazio colore determina il grado di "perdita" dei colori nell'immagine esportata. In YUV 4:4:4 non c'è perdita, mentre YUV 4:2:2 e YUV 4:1:1 rappresentano il compromesso decrescente tra dimensione del file (spazio nel disco) e qualità nella rappresentazione dei colori.

Codifica progressiva. Consente il caricamento dell'immagine per stadi. All'inizio la visualizzazione è a bassa risoluzione, quindi la qualità aumenta man mano che l'immagine continua a caricarsi.

Impostazione qualità di compressione. Consente di controllare il rapporto di compressione in funzione della qualità dell'immagine. Più alta è la qualità dell'immagine, maggiore è la dimensione del file esportato.

Operazioni colore. Sono disponibili le seguenti opzioni:

- **Inverti.** Ogni pixel viene salvato come l'inverso del colore originale.
- **Correzione gamma.** Consente di regolare l'intensità dei colori nel grafico esportato modificando la costante gamma utilizzata per creare una mappa dei valori di intensità. Essenzialmente questa funzione viene utilizzata per schiarire o scurire l'immagine bitmap. Il valore può variare da 0,10 (più scuro) a 6,5 (più chiaro).

Opzioni di esportazione dei grafici BMP e PICT

Profondità colore. Consente di determinare il numero di colori nel grafico esportato. Un grafico salvato con qualsiasi profondità di colore sarà caratterizzato dal minimo del numero di colori effettivamente utilizzati e dal massimo del numero di colori consentiti dalla profondità. Ad esempio, se il grafico contiene tre colori (rosso, bianco e nero) e viene salvato con il formato a 16 colori, il grafico rimarrà a 3 colori.

- Se il numero di colori nel grafico supera il numero di colori corrispondente alla profondità, verrà eseguita la retinatura dei colori per replicare quelli presenti nel grafico.
- Profondità schermo corrente è il numero di colori correntemente visualizzato sullo schermo del computer.

Operazioni colore. Sono disponibili le seguenti opzioni:

- **Inverti.** Ogni pixel viene salvato come l'inverso del colore originale.
- **Correzione gamma.** Consente di regolare l'intensità dei colori nel grafico esportato modificando la costante gamma utilizzata per creare una mappa dei valori di intensità. Essenzialmente questa funzione viene utilizzata per schiarire o scurire l'immagine bitmap. Il valore può variare da 0,10 (più scuro) a 6,5 (più chiaro).

Usa compressione RLE. (Solo BMP). Una tecnica di compressione senza perdita supportata dai formati di file di Windows più comuni. Una compressione senza perdita non implica una perdita di qualità dell'immagine a favore di file più piccoli.

Opzioni di esportazione dei grafici PNG e TIFF

Profondità colore. Consente di determinare il numero di colori nel grafico esportato. Un grafico salvato con qualsiasi profondità di colore sarà caratterizzato dal minimo del numero di colori effettivamente utilizzati e dal massimo del numero di colori consentiti dalla profondità. Ad esempio, se il grafico contiene tre colori (rosso, bianco e nero) e viene salvato con il formato a 16 colori, il grafico rimarrà a 3 colori.

- Se il numero di colori nel grafico supera il numero di colori corrispondente alla profondità, verrà eseguita la retinatura dei colori per replicare quelli presenti nel grafico.
- Profondità schermo corrente è il numero di colori correntemente visualizzato sullo schermo del computer.

Operazioni colore. Sono disponibili le seguenti opzioni:

- **Inverti.** Ogni pixel viene salvato come l'inverso del colore originale.
- **Correzione gamma.** Consente di regolare l'intensità dei colori nel grafico esportato modificando la costante gamma utilizzata per creare una mappa dei valori di intensità. Essenzialmente questa funzione viene utilizzata per schiarire o scurire l'immagine bitmap. Il valore può variare da 0,10 (più scuro) a 6,5 (più chiaro).

Trasparenza. Consente di selezionare un colore che diventerà trasparente nel grafico esportato. La funzione è disponibile solo con esportazione a 16,8 milioni di colori (32 bit). Immettere valori interi compresi tra 0 e 255 per ogni colore. Il valore predefinito per ciascun colore è 255, che crea un colore trasparente bianco.

Formato. (Solo TIFF). Consente di impostare lo spazio colore e comprimere il grafico esportato. Tutte le profondità di colore presenti sono disponibili con il formato RGB. Solo il formato 16,8 milioni di colori (24 e 32 bit) è disponibile in modalità CMYK. Con l'opzione YCbCr, è disponibile solo il formato 16,8 milioni di colori (24 bit).

Opzioni di esportazione dei grafici EPS

Alberi, mappe e grafici interattivi

Per gli alberi (opzione Alberi di classificazione), le mappe (opzione Mappe) ed i grafici interattivi (menu Grafici, sottomenu Interattivi), sono disponibili le seguenti opzioni EPS:

Anteprima immagine. Consente di salvare un'immagine di anteprima all'interno dell'immagine EPS. Un'immagine di anteprima viene utilizzata principalmente quando un file EPS viene posizionato all'interno di un altro documento. Molte applicazioni non possono visualizzare un'immagine EPS sullo schermo ma possono visualizzare l'anteprima salvata con l'immagine. L'anteprima dell'immagine supporta i formati WMF (più piccolo e più scalabile) e TIFF (più portatile e supportato da altre piattaforme). Controllare l'applicazione in cui si desidera includere il grafico EPS per visualizzare i formati di applicazioni supportati.

Caratteri. Controlla la modalità di gestione dei caratteri TrueType nelle immagini EPS.

- **Incorpora come TrueType nativo.** Incorpora molti dei dati dei caratteri in EPS. Il carattere PostScript risultante viene denominato carattere Type 42. *Nota*: Non tutte le stampanti PostScript possiedono driver di livello 3 che possono leggere i caratteri Type 42.
- **Converti in caratteri PostScript.** Consente di convertire i caratteri TrueType in caratteri PostScript (Type 1) basati su famiglie di caratteri. Ad esempio, Times New Roman viene convertito in Times e Arial in Helvetica. *Nota*: Questo formato non è consigliato per i grafici interattivi che utilizzano il carattere SPSS

marker (ad esempio, i grafici a dispersione), dato che non ci sono equivalenti PostScript significativi per i simboli SPSS TrueType marker.

- **Sostituisci caratteri con curve.** Consente di convertire i caratteri TrueType in dati di curva PostScript. Il testo non può più essere modificato come testo nelle applicazioni che consentono la modifica dei grafici EPS. Tale opzione comporta una perdita in termini di qualità ma è utile se si possiede una stampante PostScript che non supporta caratteri Type 42 ed è necessario preservare gli speciali simboli TrueType, come quelli utilizzati nei grafici a dispersione interattivi.

Altri grafici

Per tutti gli altri grafici, sono disponibili le seguenti opzioni EPS:

Includi anteprima TIFF. Salva un'anteprima con l'immagine EPS in formato TIFF in modo che possa essere visualizzata in applicazioni che non supportano la visualizzazione delle immagini EPS sullo schermo.

Caratteri. Controlla la modalità di gestione dei caratteri nelle immagini EPS.

- **Sostituisci caratteri con curve.** Consente di convertire i caratteri in dati di curva PostScript. Il testo non può più essere modificato come testo nelle applicazioni che consentono la modifica dei grafici EPS. Questa opzione è utile se i caratteri usati nel grafico non sono disponibili sulla periferica di output.
- **Usa caratteri di riferimento.** Se i caratteri usati nel grafico sono disponibili per la periferica di output, vengono utilizzati questi caratteri. Altrimenti, la periferica di output utilizza altri caratteri.

Opzioni di esportazione dei grafici WMF

Aldus placeable. Offre un grado di indipendenza dalla periferica (stessa dimensione fisica se aperto a 96 o 120 dpi), ma non tutte le applicazioni supportano questo formato.

Windows Standard. Supportato dalla maggior parte delle applicazioni che possono visualizzare i metafile di Windows.

Per impostare le opzioni di esportazione dei grafici

- Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- Dai menu, scegliere:
  File
  Esporta...
- Per l'output, fare clic su Opzioni, selezionare il formato di esportazione e fare clic su Opzioni.
- Per l'opzione Solo grafici, selezionare il formato di esportazione e fare clic su Opzioni.

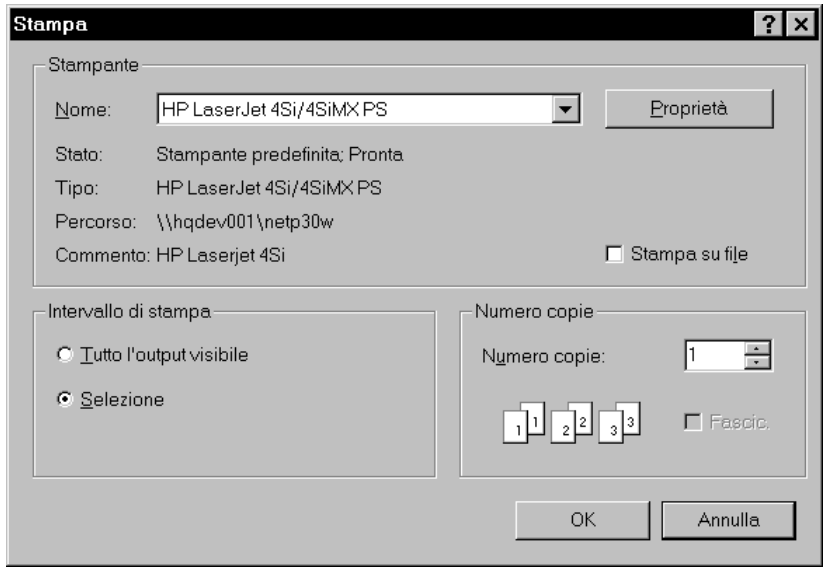
Stampa di documenti del Viewer

Sono disponibili due opzioni per stampare il contenuto della finestra Viewer:

Tutto l'output visibile. Consente di stampare solo gli elementi visualizzati. Gli elementi nascosti, ovvero contrassegnati da un'icona a forma di libro chiuso o nascosti negli strati compressi nella vista riassuntiva, non verranno stampati.

Selezione. Consente di stampare solo gli elementi selezionati nella vista riassuntiva e/o nell'output.

Figura 10-9
Finestra di dialogo della stampa del Viewer



Per stampare output e grafici

- ▶ Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- ▶ Dai menu, scegliere:

  File
  Stampa...
- ▶ Selezionare le impostazioni di stampa desiderate.
- ▶ Fare clic su OK per stampare.

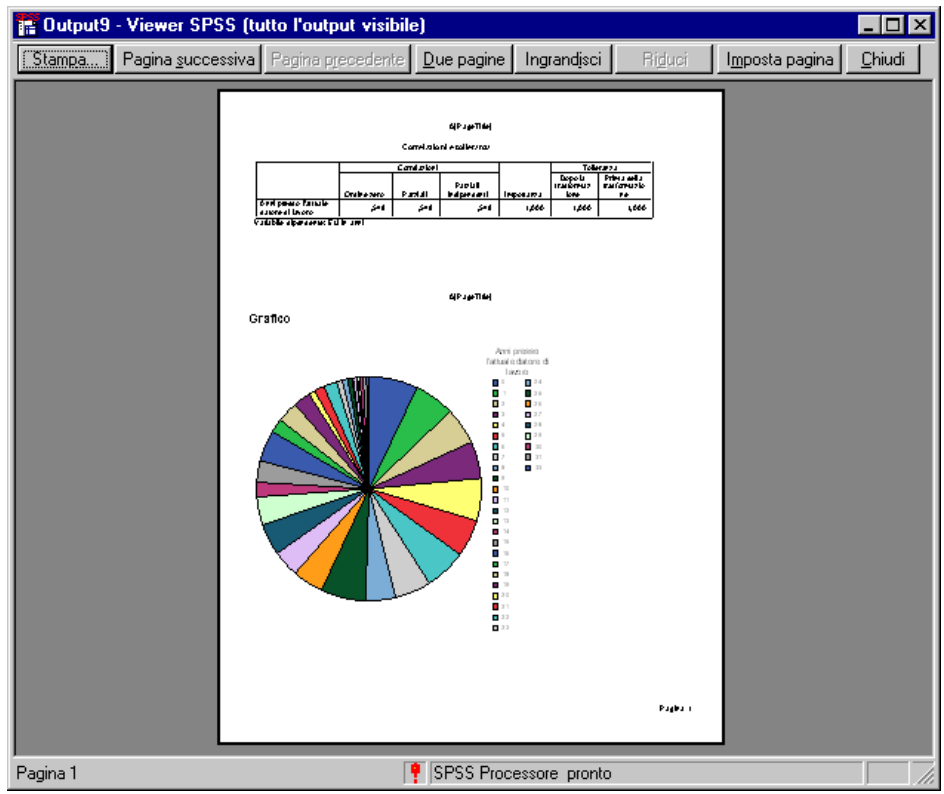
Anteprima di stampa

Nell'Anteprima di stampa è possibile visualizzare il contenuto che verrà stampato su ogni pagina dei documenti del Viewer. Prima di stampare un documento del Viewer è consigliabile attivare l'Anteprima di stampa, in quanto vi vengono visualizzati elementi che potrebbero non risultare visibili nel Viewer, ad esempio:

- ■ Interruzioni di pagina

- Strati nascosti di tabelle pivot
- Interruzioni in tabelle di grandi dimensioni
- Output completo di tabelle di grandi dimensioni
- Intestazioni e piè di pagina stampati in ciascuna pagina

Figura 10-10
Anteprima di stampa



Se nel Viewer è selezionato output, nell'anteprima verrà visualizzata solo la selezione attiva. Per visualizzare in anteprima tutto l'output è necessario verificare che nel Viewer non sia stata effettuata alcuna selezione.

Per visualizzare l'anteprima di stampa

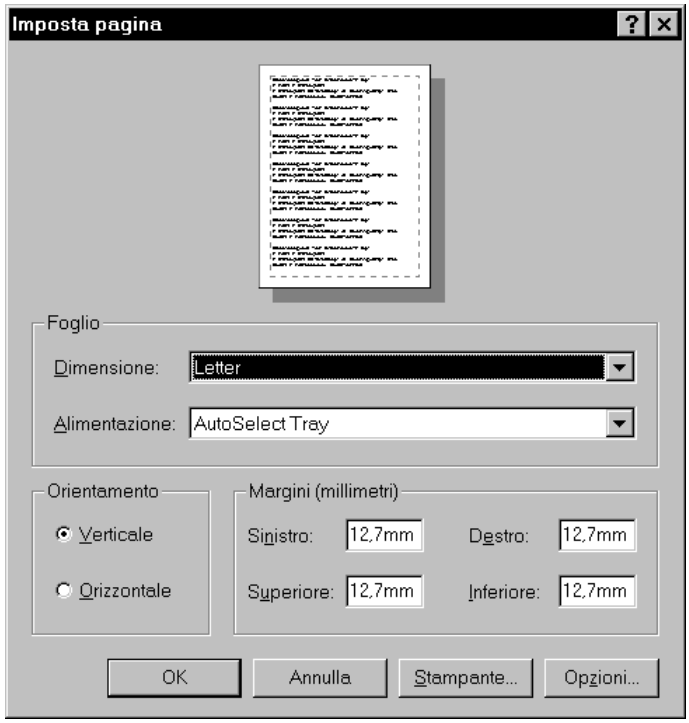
- ▶ Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- ▶ Dai menu, scegliere:
  File
  Anteprima di stampa

Imposta pagina

Nella finestra di dialogo Imposta pagina è possibile impostare:

- Le dimensioni e l'orientamento della carta
- I margini della pagina
- Le intestazioni e i piè di pagina
- La numerazione delle pagine
- Le dimensioni di stampa dei grafici

Figura 10-11
Finestra di dialogo Imposta pagina



Le impostazioni della finestra di dialogo Imposta pagina vengono salvate insieme ai documenti del Viewer. Le opzioni Imposta pagina influiscono solo sulle impostazioni per la stampa dei documenti del Viewer. Tali impostazioni non producono alcun effetto sulla stampa dei dati dall'Editor dei dati o della sintassi da una finestra di sintassi.

Per modificare l'impostazione della pagina

- Attivare la finestra del Viewer (fare clic su un punto qualsiasi della finestra).
- Dai menu, scegliere:

  File
  Imposta pagina...

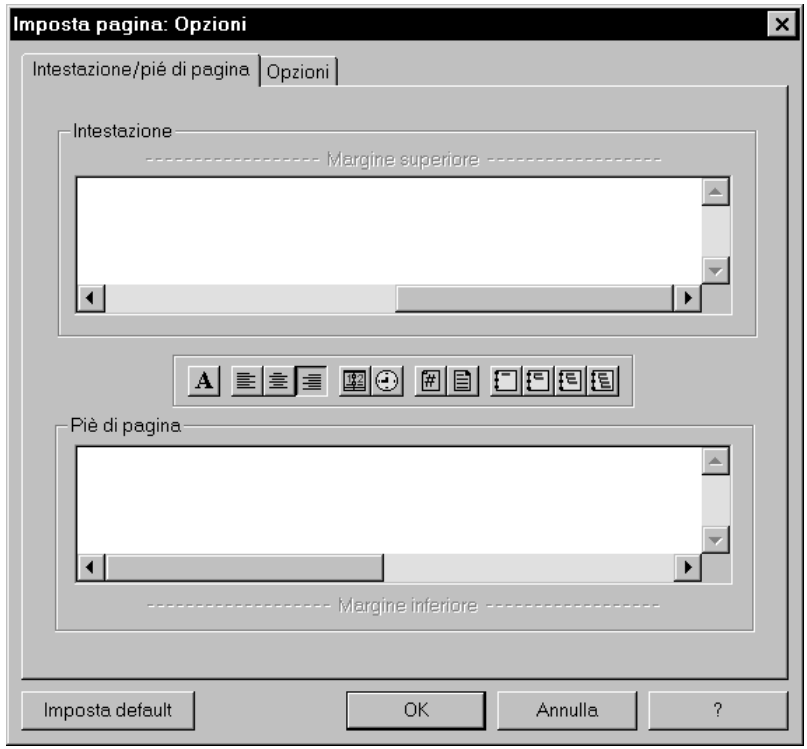
- Modificare le impostazioni e quindi fare clic su OK.

Imposta pagina: intestazioni e piè di pagina

Le intestazioni e i piè di pagina sono le informazioni che vengono stampate all'inizio e alla fine di ogni pagina. È possibile specificare qualsiasi testo che si desidera utilizzare come intestazione o piè di pagina. È inoltre possibile utilizzare la barra degli strumenti al centro della finestra di dialogo per inserire:

- Data e orario
- Numeri di pagina
- Nome file Viewer
- Etichette dell'intestazione
- Titoli e sottotitoli di pagina

Figura 10-12
Imposta pagina: Opzioni, scheda Intestazione/Piè di pagina



Le etichette delle intestazioni indicano l'intestazione di primo, secondo, terzo e/o quarto livello del primo elemento di ciascuna pagina.

Vengono stampati i titoli e i sottotitoli della pagina corrente. È possibile creare titoli e sottotitoli di pagina scegliendo il comando Nuovo titolo di pagina dal menu Inserisci del Viewer oppure con i comandi `TITLE` e `SUBTITLE` della sintassi. Se non sono stati specificati titoli o sottotitoli, questa impostazione verrà ignorata.

Nota: le caratteristiche del carattere di titoli e sottotitoli della nuova pagina possono essere definite utilizzando la scheda Viewer della finestra di dialogo Opzioni (accessibile selezionando Opzioni nel menu Modifica). Le impostazioni dei caratteri di titoli e sottotitoli esistenti possono essere ridefinite modificando i titoli nel Viewer.

Per visualizzare in anteprima titoli e note a pié di pagina, selezionare Anteprima di stampa dal menu File.

Imposta pagina: Opzioni

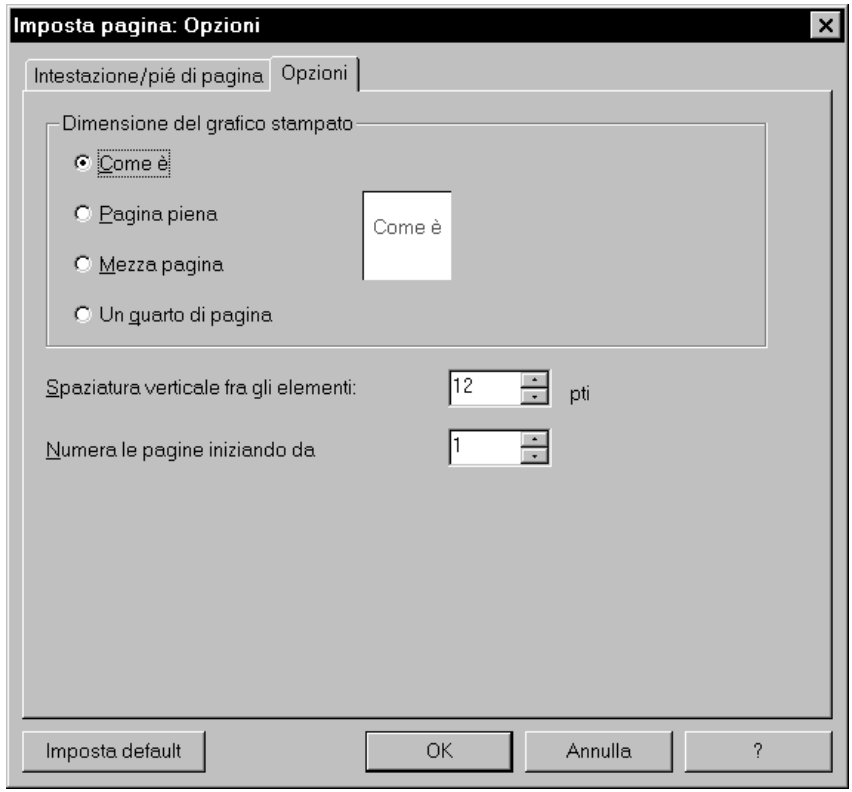
In questa finestra di dialogo è possibile impostare le dimensioni del grafico stampato, lo spazio tra gli elementi dell'output stampato e la numerazione delle pagine.

Dimensione del grafico stampato. Consente di impostare le dimensioni del grafico stampato in relazione a quelle definite per la pagina. Le dimensioni del grafico stampato non influiscono sul rapporto larghezza/altezza del grafico stesso. Le dimensioni totali di un grafico stampato sono limitate sia dall'altezza che dalla larghezza. Quando i bordi esterni del grafico raggiungono il bordo destro e sinistro della pagina, non è possibile aumentare ulteriormente le dimensioni del grafico per occupare in altezza la rimanente parte della pagina.

Spaziatura verticale fra gli elementi. Consente di impostare lo spazio verticale fra gli elementi stampati. Ogni tabella pivot, grafico e oggetto di testo rappresenta un elemento distinto. Questa impostazione non influisce sulla visualizzazione degli elementi nel Viewer.

Numera le pagine iniziando da. Le pagine vengono numerate in sequenza, iniziando dal numero indicato.

Figura 10-13
Imposta pagina: Opzioni, scheda Opzioni



Salvataggio dell'output

È possibile salvare il contenuto del Viewer come documento. Il documento salvato include entrambi i riquadri della finestra del Viewer, ovvero la vista riassuntiva e il contenuto.

Per salvare un documento Viewer

- Dai menu del Viewer, scegliere:

File
Salva

- Digitare il nome del documento, quindi fare clic su Salva.

  Per salvare l'output in formati esterni (ad esempio, HTML o di testo), utilizzare il comando Esporta del menu File Questa funzione non è disponibile nella versione autonoma di SmartViewer.

Opzione Salva con password

Questa opzione consente di proteggere i file del Viewer tramite una password.

Password. La password può essere scritta in lettere maiuscole e minuscole e può contenere fino a 16 caratteri. Se si assegna una password, questa deve sempre essere immessa per visualizzare il file.

Codice OEM. Lasciare vuoto questo campo se non si è in possesso di un contratto stipulato con SPSS Inc. e relativo alla ridistribuzione di SmartViewer. Il codice di licenza OEM è incluso nel contratto.

Per salvare i file del Viewer con una password

- Dai menu del Viewer, scegliere:
  File
  Salva con password...
- Immettere la password.
- Immettere nuovamente la password per confermarla, quindi fare clic su OK.
- Immettere un nome di file nella finestra di dialogo Salva con nome.
- Scegliere Salva.

Nota: lasciare il campo Codice OEM vuoto se non si è in possesso di un contratto stipulato con SPSS Inc. e relativo alla ridistribuzione di SmartViewer.

Capitolo
11

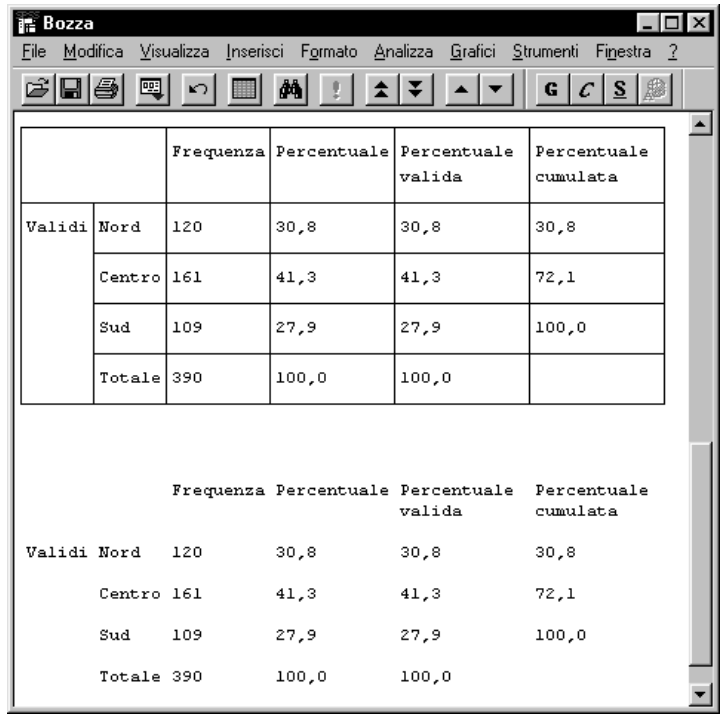
Bozza

La funzione Bozza consente di ottenere risultati in forma di bozza, ad esempio:

- Output testuale semplice (anziché tabelle pivot)
- Grafici come immagini metafile (anziché oggetti grafici)

L'output testuale della Bozza può essere modificato, i grafici possono essere ridimensionati ed entrambi possono essere incollati in altre applicazioni. Tuttavia, i grafici non possono essere modificati e le funzioni interattive delle tabelle pivot e dei grafici non sono disponibili.

Figura 11-1
Finestra Bozza



		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Nord	120	30,8	30,8	30,8
	Centro	161	41,3	41,3	72,1
	Sud	109	27,9	27,9	100,0
	Totale	390	100,0	100,0	

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Nord	120	30,8	30,8	30,8
	Centro	161	41,3	41,3	72,1
	Sud	109	27,9	27,9	100,0
	Totale	390	100,0	100,0	

Per creare output della Bozza

- Dai menu, scegliere:

  File
  Nuovo
  Output bozza

- Per impostare l'output della Bozza come tipo di output di default, dai menu scegliere:

  Modifica
  Opzioni...

- Fare clic sulla scheda Generale.

- Selezionare Bozza in Tipo di Viewer all'avvio.

Nota: Il nuovo output viene sempre visualizzato nella finestra del Viewer designata. Se sono aperte una finestra del Viewer e una finestra Bozza, la **finestra designata** sarà l'ultima aperta oppure quella specificata mediante il pulsante Designa finestra (il punto esclamativo) della barra degli strumenti.

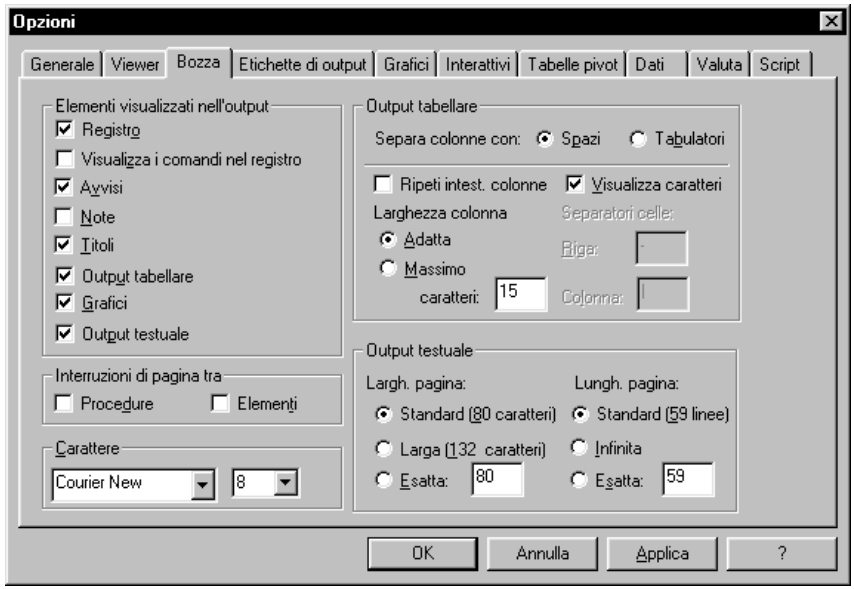
Controllo del formato dell'output bozza

L'output visualizzato come tabella pivot nel Viewer viene convertito in output testuale nella Bozza. Le impostazioni predefinite per la conversione dell'output della tabella pivot includono quanto segue:

- La larghezza di ciascuna colonna è impostata sulla lunghezza della relativa etichetta. Le etichette sono contenute in un'unica riga.
- L'allineamento viene controllato dagli spazi (e non dalle tabulazioni).
- Come separatori di riga e di colonna vengono impiegate le linee continue del carattere SPSS Marker Set.
- Se le linee continue non sono disponibili, come separatori di colonna vengono utilizzati i caratteri di linea verticale (|) e come separatori di riga vengono utilizzati i trattini (–).

È possibile controllare il formato del nuovo output della bozza utilizzando le opzioni della Bozza (menu Modifica, Opzioni, scheda Bozza).

Figura 11-2
Opzioni Output bozza



Larghezza colonna. Per limitare la larghezza delle tabelle contenenti etichette lunghe, selezionare il numero massimo di caratteri per la larghezza della colonna. Le etichette più lunghe della larghezza specificata vengono mandate a capo per rispettare la larghezza massima.

Figura 11-3
Output bozza prima e dopo aver impostato la larghezza massima di colonna

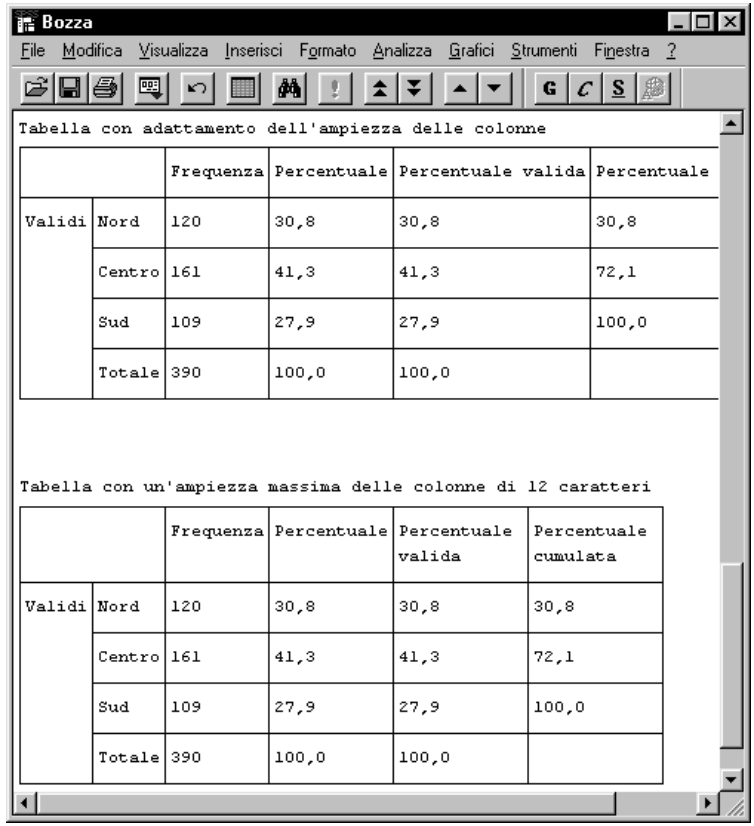


Tabella con adattamento dell'ampiezza delle colonne

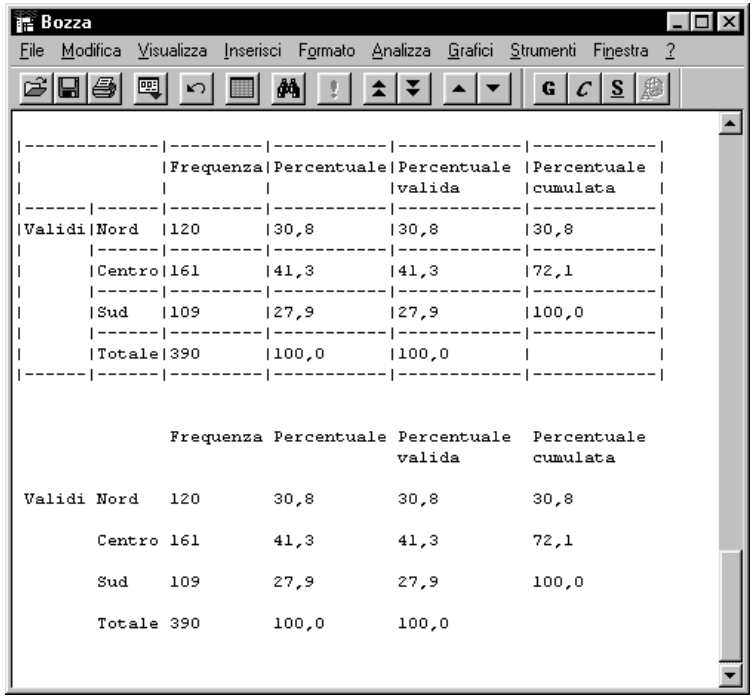
		Frequenza	Percentuale	Percentuale valida	Percentuale
Validi	Nord	120	30,8	30,8	30,8
	Centro	161	41,3	41,3	72,1
	Sud	109	27,9	27,9	100,0
	Totale	390	100,0	100,0	

Tabella con un'ampiezza massima delle colonne di 12 caratteri

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Nord	120	30,8	30,8	30,8
	Centro	161	41,3	41,3	72,1
	Sud	109	27,9	27,9	100,0
	Totale	390	100,0	100,0	

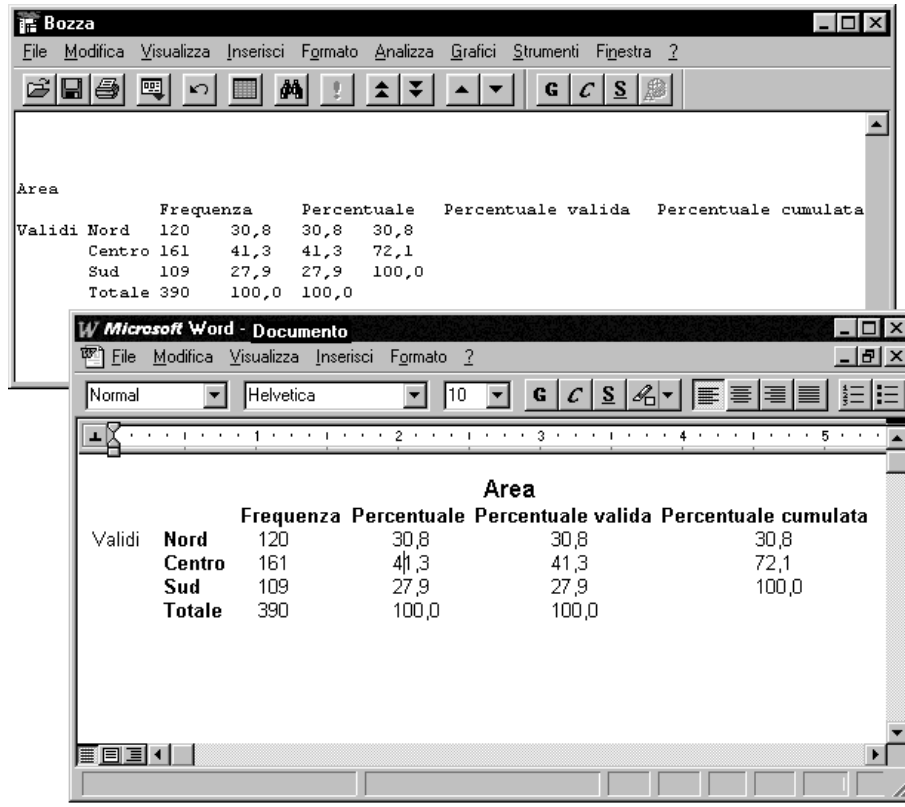
Separatori di riga e di colonna. Per gestire i separatori di righe e colonne visualizzati nell'output della bozza, in alternativa ai caratteri di linea continua è possibile utilizzare le impostazioni dei separatori di celle. È possibile specificare separatori di celle diversi o immettere spazi se non si vogliono utilizzare caratteri per contrassegnare le righe e le colonne. Per impostare i separatori di celle, è necessario deselezionare Visualizza caratteri.

Figura 11-4
Output bozza prima e dopo aver impostato i separatori di cella


Bozza
File Modifica Visualizza Inserisci Formato Analizza Grafici Strumenti Finestra ?
|---|---|---|---|---|---|
		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Nord	120	30,8	30,8	30,8
	Centro	161	41,3	41,3	72,1
	Sud	109	27,9	27,9	100,0
	Totale	390	100,0	100,0	
Frequenza Percentuale Percentuale valida Percentuale cumulata
Validi Nord 120 30,8 30,8 30,8
Centro 161 41,3 41,3 72,1
Sud 109 27,9 27,9 100,0
Totale 390 100,0 100,0

Colonne separate da spazi e da tabulatori. La Bozza è stata studiata per visualizzare output separati da spazi con un carattere a spaziatura fissa (proporzionale). Per incollare un output bozza in un'altra applicazione è necessario utilizzare un carattere a spaziatura fissa per allineare in modo appropriato le colonne separate da spazi. Se si seleziona Tabulatori come separatore di colonna, è possibile utilizzare qualsiasi carattere dell'altra applicazione e impostare i tabulatori in modo da allineare l'output in modo corretto. Un output separato da tabulatori non verrà tuttavia allineato in modo corretto nella Bozza.

Figura 11-5
Output separato da tabulatori nella Bozza e formattato con un elaboratore di testi.


Bozza
File Modifica Visualizza Inserisci Formato Analizza Grafici Strumenti Finestra ?
Area
Frequenza Percentuale Percentuale valida Percentuale cumulata
Validi Nord 120 30,8 30,8 30,8
Centro 161 41,3 41,3 72,1
Sud 109 27,9 27,9 100,0
Totale 390 100,0 100,0
Microsoft Word - Documento
File Modifica Visualizza Inserisci Formato ?
Normal
Helvetica
10
Area
Frequenza Percentuale Percentuale valida Percentuale cumulata
Validi Nord 120 30,8 30,8 30,8
Centro 161 41,3 41,3 72,1
Sud 109 27,9 27,9 100,0
Totale 390 100,0 100,0

Per impostare le opzioni di Bozza

- Dai menu, scegliere:
  Modifica
  Opzioni...
- Selezionare la scheda Bozza.
- Selezionare le opzioni desiderate.
- Fare clic su OK o su Applica.

Le opzioni di visualizzazione dell'output di Bozza influiscono solo sul nuovo output prodotto dopo che le impostazioni sono state modificate. Le modifiche di tali impostazioni non influiscono sull'output già visualizzato in Bozza.

Caratteri in Bozza

È possibile modificare gli attributi dei caratteri dell'output testuale della Bozza (carattere, dimensione, stile). Se tuttavia si utilizzano i caratteri di linea continua per i bordi di righe e colonne, è possibile ottenere un allineamento esatto delle colonne con testo separato da spazi soltanto utilizzando un tipo di carattere proporzionale, ad esempio Courier. L'allineamento della colonna può inoltre essere influenzato da altre modifiche apportate ai caratteri, ad esempio modifiche relative alla dimensione e allo stile (grassetto e corsivo) e applicate solo a una parte della tabella.

Bordi di righe e colonne. Per i bordi predefiniti di righe e colonne a linee continue viene utilizzato il carattere SPSS Marker Set. Questi caratteri di linea continua per il disegno dei bordi non sono supportati da altri caratteri.

Per modificare il tipo di caratteri nella Bozza

- Selezionare il testo a cui applicare la modifica del carattere.
- Dai menu di Bozza scegliere:

  Formato
  Carattere...

- Selezionare gli attributi di carattere da applicare al testo selezionato.

Per stampare l'output della Bozza

- Dai menu di Bozza scegliere:

  File
  Stampa...

Per stampare la parte selezionata dell'output bozza:

- Selezionare l'output da stampare.

▶ Dai menu, scegliere:
File
Stampa...

▶ Scegliere Selezione.

Anteprima di stampa della Bozza

Nell'anteprima di stampa viene visualizzato il contenuto dei documenti di Bozza che verrà stampato su ciascuna pagina. Prima di stampare un documento di Bozza è consigliabile attivare l'Anteprima di stampa, in quanto vi vengono visualizzati elementi che potrebbero risultare di dimensioni superiori alla pagina, ad esempio:

- Tabelle lunghe
- Tabelle di grandi dimensioni prodotte dalla conversione dell'output di tabelle pivot senza il controllo della larghezza di colonna
- Output testuale creato con l'opzione di larghezza di pagina Larga (opzioni Bozza) e l'orientamento di stampa Verticale

Se è troppo grande per la pagina, l'output viene troncato e non viene stampato su un'altra pagina. Per impedire che ciò avvenga è possibile procedere in diversi modi:

- Utilizzare un carattere di dimensioni minori (menu Formato, Carattere).
- Selezionare l'orientamento Orizzontale (menu File, Imposta pagina).
- In caso di nuovo output, ridurre la larghezza massima di colonna (menu Modifica, Opzioni, scheda Bozza).

In caso di tabelle lunghe, utilizzare le interruzioni di pagina (menu Inserisci, Interruzione di pagina) per definire le interruzioni di tabella fra le pagine.

Per visualizzare l'anteprima di stampa della Bozza

▶ Dai menu di Bozza scegliere:
File
Anteprima di stampa

Per salvare l'output della Bozza come testo

▶ Dai menu di Bozza scegliere:

File
  Salva

L'output della Bozza viene salvato in formato RTF (Rich Text Format).

Per salvare l'output Bozza come testo

▶ Dai menu di Bozza scegliere:

File
  Esporta...

È possibile esportare tutto il testo o solo la parte selezionata. Nei file esportati viene salvato solo l'output testuale (output della tabella pivot e output testuale convertito), mentre i grafici non vengono inclusi.

Capitolo
12

Tabelle pivot

Nel Viewer molti risultati sono presentati in tabelle che possono essere modificate in modo interattivo. In altri termini, è possibile ridisporre le righe, le colonne e gli strati.

Gestione di una tabella pivot

Sono disponibili le seguenti opzioni per la gestione delle tabelle pivot:

- Trasposizione di righe e colonne
- Spostamento di righe e colonne
- Creazione di strati multidimensionali
- Raggruppamento e separazione di righe e colonne
- Possibilità di visualizzare e nascondere le celle
- Rotazione delle etichette delle righe e delle colonne
- Ricerca di definizioni di termini

Per modificare una tabella pivot

▶ Fare doppio clic sulla tabella.

Verrà attivato l'Editor di tabelle pivot.

Per modificare due o più tabelle pivot alla volta

▶ Fare clic con il pulsante destro del mouse sulla tabella pivot.

▶ Dal menu di scelta rapida, scegliere:

Oggetto tabella pivot SPSS
Apri

- Ripetere l'operazione per ciascuna tabella pivot che si desidera modificare.

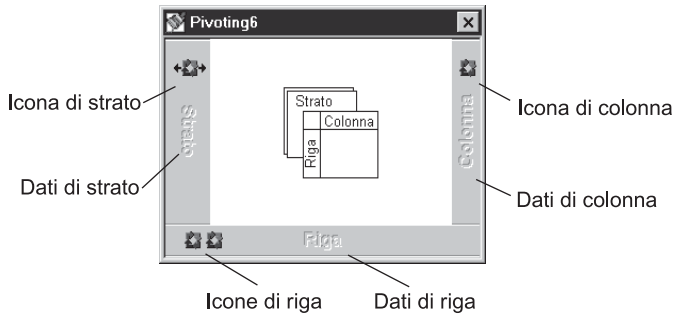
  Ciascuna tabella pivot verrà visualizzata in una finestra distinta per la modifica.

Per effettuare il pivoting di una tabella utilizzando le icone dei dati

- Attivare la tabella pivot.
- Dai menu della tabella pivot, scegliere:

  Pivot
  Pivoting
- Posizionare il puntatore sopra ciascuna icona per visualizzare il suggerimento che fornisce informazioni sulla dimensione tabella rappresentata dall'icona.
- Trascinare un'icona tra i dati.

Figura 12-1
Pivoting



In questo modo la tabella viene riorganizzata. Si supponga, ad esempio, che l'icona rappresenti una variabile con le categorie Sì e No. Se si trascina l'icona dall'area Riga all'area Colonna, Sì e No diventano etichette di colonne. Prima dello spostamento, Sì e No erano etichette di riga.

Per identificare le dimensioni di una tabella pivot

- Attivare la tabella pivot.
- Se il pivoting non è stato attivato, dai menu della tabella pivot, scegliere:

  Pivot
  Pivoting

- Fare clic su un'icona e tenere premuto il pulsante del mouse.

  Verranno evidenziate le etichette delle dimensioni nella tabella pivot.

Per trasporre righe e colonne

- Dai menu della tabella pivot, scegliere:
  Pivot
  Trasponi righe e colonne

  Questa operazione equivale a trascinare tutte le icone di riga nei dati di colonna e tutte le icone di colonna nei dati di riga.

Per modificare l'ordine di visualizzazione

L'ordine delle icone in una dimensione riflette l'ordine di visualizzazione degli elementi nella tabella pivot. Per modificare l'ordine di visualizzazione degli elementi in una dimensione:

- Attivare la tabella pivot.
- Se il pivoting non è stato attivato, dai menu della tabella pivot, scegliere:
  Pivot
  Pivoting
- Trascinare le icone in ciascuna dimensione nell'ordine desiderato (da sinistra a destra o dall'alto in basso).

Per spostare righe e colonne in una tabella pivot

- Attivare la tabella pivot.
- Fare clic sull'etichetta della riga o della colonna che si desidera spostare.
- Trascinare l'etichetta nella nuova posizione.
- Dal menu di scelta rapida selezionare Inserisci prima o Scambia.

Nota: verificare che il comando Trascina per copiare del menu Modifica *non* sia attivato. Se il comando Trascina per copiare è attivato, deselezionarlo.

Per raggruppare righe e colonne e inserire etichette di gruppo

- ▶ Attivare la tabella pivot.
- ▶ Selezionare le etichette per le righe o le colonne che si desidera raggruppare. Fare clic e trascinare oppure fare clic tenendo premuto Maiusc per selezionare più etichette.
- ▶ Dai menu, scegliere:

  Modifica
  Gruppo

  Verrà inserita automaticamente un'etichetta di gruppo. Fare doppio clic sull'etichetta di gruppo per modificarne il testo.

Figura 12-2
Gruppi ed etichette di righe e colonne

		Etichetta Gruppo di colonne		
		Fammina	Maschio	Totale
Ettichetta Gruppo di righe	Clerical	206	157	363
	Custodial		27	27
	Manager	10	74	84

Nota: per aggiungere righe o colonne a un gruppo esistente, è innanzitutto necessario separare gli elementi costitutivi del gruppo e quindi creare un nuovo gruppo che includa elementi aggiuntivi.

Per separare righe o colonne e rimuovere etichette di gruppo

- ▶ Attivare la tabella pivot.
- ▶ Fare clic su qualunque punto all'interno dell'etichetta del gruppo per le righe e colonne da rimuovere dal gruppo.
- ▶ Dai menu, scegliere:

  Modifica
  Annulla raggruppamento

  Annullando il raggruppamento verrà eliminata automaticamente l'etichetta di gruppo.

Per ruotare le etichette delle tabelle pivot

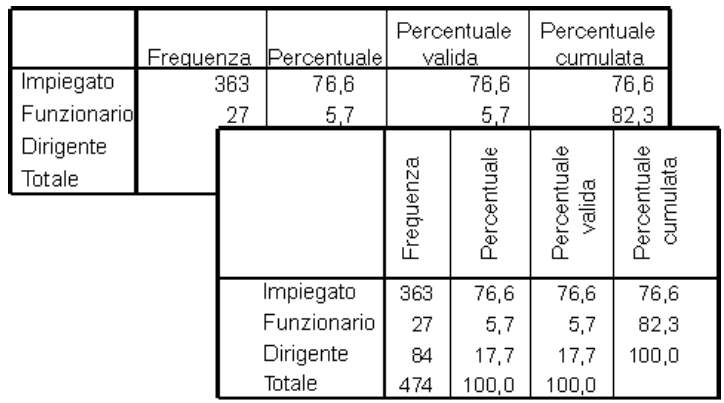
- ▶ Attivare la tabella pivot.
- ▶ Dai menu, scegliere:

  Formato
  Ruota etichette colonne interne

  oppure

  Ruota etichette righe esterne

Figura 12-3
Etichette di colonna ruotate



	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Impiegato	363	76,6	76,6	76,6
Funzionario	27	5,7	5,7	82,3
Dirigente				
Totale				

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Impiegato	363	76,6	76,6	76,6
Funzionario	27	5,7	5,7	82,3
Dirigente	84	17,7	17,7	100,0
Totale	474	100,0	100,0	

È possibile ruotare solo le etichette delle colonne interne e le etichette delle righe esterne.

Per ripristinare le impostazioni predefinite delle tabelle pivot

Dopo aver eseguito una o più operazioni di pivoting, è possibile tornare alla disposizione originale della tabella pivot.

- ▶ Dal menu Pivot scegliere Ripristina pivot di default.

Verranno ripristinate solo le modifiche risultanti dal pivoting di righe, colonne ed elementi dello strato tra dimensioni. L'operazione non produce alcun effetto su modifiche quali il raggruppamento o la separazione, né sullo spostamento di righe o colonne.

Per trovare la definizione di un'etichetta in una tabella pivot

È possibile visualizzare l'Aiuto contestuale relativo alle etichette delle celle nelle tabelle pivot. Se, ad esempio, viene visualizzata l'etichetta *Media*, è possibile ottenere una definizione della media.

- Fare clic con il pulsante destro del mouse su una cella dell'etichetta.
- Dal menu di scelta rapida scegliere Guida rapida.

  È necessario fare clic con il pulsante destro del mouse sulla cella dell'etichetta e non sulle celle di dati incluse nella riga o nella colonna. L'Aiuto contestuale non è disponibile per le etichette definite dall'utente, ad esempio nomi di variabili o etichette dei valori.

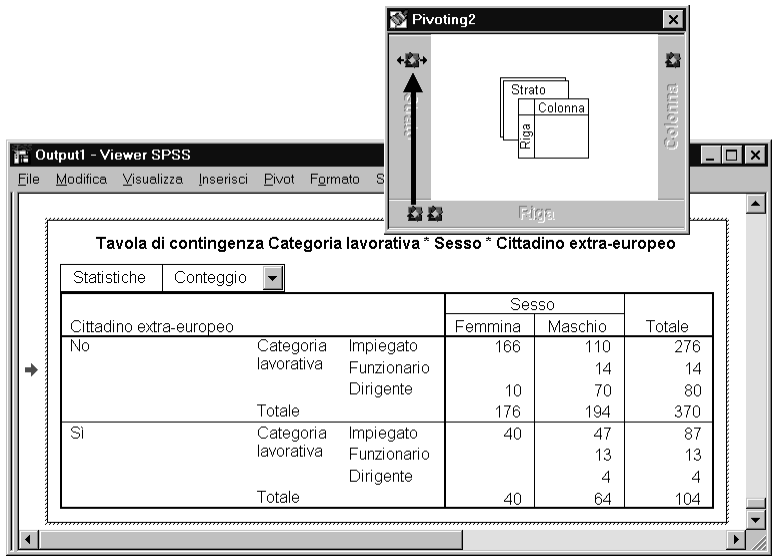
Utilizzo degli strati

È possibile visualizzare una tabella bidimensionale distinta per ciascuna categoria o combinazione di categorie. La tabella può essere considerata come formata da diversi strati, di cui solo il primo è visibile.

Per creare e visualizzare gli strati

- Attivare la tabella pivot.
- >> Nel menu Pivot selezionare Pivoting (se l'opzione non è ancora stata selezionata).
- Trascinare un'icona dai dati di riga o di colonna ai dati di strato.

Figura 12-4
Spostamento di categorie negli strati



All'icona di ogni strato corrispondono le frecce destra e sinistra. La tabella visibile corrisponde allo strato superiore.

Figura 12-5
Categorie in strati distinti

Conteggio

Cittadino extra-europeo: Sì

		Sesso		
		Femmina	Maschio	Totale
Categoria lavorativa	Impiegato	40	47	87
	Funzionario		13	13
			4	4

Cittadino extra-europeo: No

		Sesso		
		Femmina	Maschio	Totale
Categoria lavorativa	Impiegato	166	110	276
	Funzionario	0	14	14
	Dirigente	10	70	80
Totale		176	194	370

Per modificare gli strati

- Fare clic su una delle frecce dell'icona dello strato.

  oppure

- Selezionare una categoria dall'elenco di strati.

Figura 12-6
Selezione di strati dagli elenchi a discesa

Report stratificati

Area: Sud (Centro, Sud, Totale)

	N	Media	Deviazione std.	% della somma totale	% del numero di casi totale
Divisione	161	1,52	,50	40,8%	41,3%

Vai allo strato

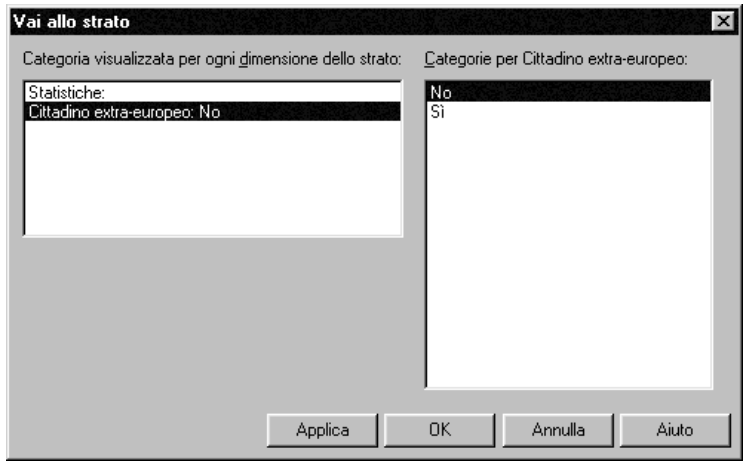
L'opzione Vai allo strato consente di modificare gli strati di una tabella pivot. Questa finestra di dialogo risulta particolarmente utile se è presente un grande numero di strati o se uno strato include numerose categorie.

Per passare a uno strato della tabella

- Dai menu della tabella pivot, scegliere:

  Pivot
  Vai allo strato...

Figura 12-7
Finestra di dialogo Vai allo strato



- Nell'elenco Categoria visibile, selezionare la dimensione di uno strato. Nell'elenco Categorie verranno visualizzate tutte le categorie per la dimensione selezionata.

- Nell'elenco Categorie, selezionare la categoria desiderata, quindi fare clic su OK. Lo strato verrà modificato e la finestra di dialogo verrà chiusa.

  Per visualizzare uno strato senza chiudere la finestra di dialogo:

- Selezionare la categoria e fare clic su Applica.

Per spostare gli strati in righe o colonne

Se la tabella visualizzata è formata da strati di cui solo il primo è visibile, è possibile visualizzare contemporaneamente tutti gli strati, sia lungo le righe che lungo le colonne. Nei dati di strato deve essere presente almeno un'icona.

- Scegliere Sposta strati nelle righe dal menu Pivot.

  oppure

- Scegliere Sposta strati nelle colonne dal menu Pivot.

  Per spostare strati nelle righe o nelle colonne è inoltre possibile trascinarne le icone sulle dimensioni degli strati, delle righe o delle colonne.

Segnalibri

I segnalibri consentono di salvare diverse visualizzazioni di una tabella pivot. I segnalibri memorizzano:

- La posizione degli elementi nelle dimensioni di riga, colonna e strato.
- L'ordine di visualizzazione degli elementi in ciascuna dimensione.
- Lo strato correntemente visualizzato per ciascun elemento dello strato.

Per ripristinare le visualizzazioni delle tabelle pivot con i relativi segnalibri

- Attivare la tabella pivot.
- Passare alla visualizzazione della tabella pivot per la quale si desidera creare un segnalibro.
- Dai menu, scegliere:
  Pivot
  Segnalibri…
- Specificare un nome per il segnalibro. Per i nomi di segnalibri non è attiva la distinzione tra maiuscole e minuscole.
- Fare clic su Aggiungi.

Ciascuna tabella pivot ha un insieme specifico di segnalibri. I nomi dei singoli segnalibri devono essere univoci all'interno di una tabella pivot, ma è possibile utilizzare nomi uguali in tabelle pivot diverse.

Per visualizzare le tabelle pivot con i relativi segnalibri

- ▶ Attivare la tabella pivot.
- ▶ Dai menu, scegliere:
  Pivot
  Segnalibri…
- ▶ Fare clic sul nome del segnalibro nell'elenco.
- ▶ Fare clic su Vai a.

Per rinominare il segnalibro di una tabella pivot

- ▶ Attivare la tabella pivot.
- ▶ Dai menu, scegliere:
  Pivot
  Segnalibri…
- ▶ Fare clic sul nome del segnalibro nell'elenco.
- ▶ Fare clic su Rinomina.
- ▶ Specificare il nuovo nome del segnalibro.
- ▶ Fare clic su OK.

Per visualizzare e nascondere le celle

È possibile nascondere molti tipi di cella, compresi:

- Etichette delle dimensioni.
- Categorie, incluse le celle dell'etichetta e dei dati di una riga o colonna.
- Etichette delle categorie, senza nascondere le celle dei dati.
- Note a piè di pagina, titoli e didascalie.

Per nascondere le righe e le colonne di una tabella

- Tenere premuti Ctrl e Alt e fare clic sull'etichetta di categoria della riga o colonna da nascondere.
- Dai menu della tabella pivot, scegliere:

  Visualizza
  Nascondi

  oppure
- Fare clic con il pulsante destro del mouse sulla riga o la colonna.
- Scegliere Nascondi categoria dal menu di scelta rapida.

Per visualizzare righe e colonne nascoste in una tabella

- Selezionare un'altra etichetta nella stessa dimensione della riga o della colonna nascosta.

  Se, ad esempio, la categoria *Femmina* della dimensione Sesso è nascosta, fare clic sulla categoria *Maschio*.
- Dai menu della tabella pivot, scegliere:

  Visualizza
  Mostra tutte le categorie in nome dimensione

  Ad esempio, scegliere Mostra tutte le categorie in Sesso.

  oppure
- Dai menu della tabella pivot, scegliere:

  Visualizza
  Mostra tutto

  Verranno visualizzate tutte le celle nascoste della tabella. Se è selezionata l'opzione Nascondi righe e colonne vuote nella finestra di dialogo Proprietà tabella per questa tabella, le righe o le colonne completamente vuote rimarranno nascoste.

Per nascondere o mostrare le etichette delle dimensioni

- Attivare la tabella pivot.

- ▶ Selezionare l'etichetta di dimensione o qualsiasi etichetta di categoria all'interno della dimensione.
- ▶ Dai menu, scegliere:
  Visualizza
  Nascondi/mostra etichetta dimensione

Per nascondere o visualizzare una nota a piè di pagina in una tabella

- ▶ Selezionare una nota a piè di pagina.
- ▶ Dai menu, scegliere:
  Visualizza
  Nascondi (o Visualizza)

Per nascondere o visualizzare una didascalia o un titolo in una tabella

- ▶ Selezionare una didascalia o un titolo.
- ▶ Dai menu, scegliere:
  Visualizza
  Nascondi (o Visualizza)

Modifica dei risultati

È possibile modificare l'aspetto e il contenuto di ciascuna tabella o elemento di output testuale. È possibile:

- Applicare un TableLook.
- Modificare le proprietà della tabella corrente.
- Modificare le proprietà delle celle della tabella.
- Modificare il testo.
- Aggiungere note a piè di pagina e didascalie alle tabelle.
- Aggiungere elementi nel Viewer
- Copiare e incollare l'output in altre applicazioni.
- Modificare le larghezze delle celle dati.

Modifica dell'aspetto delle tabelle

È possibile modificare l'aspetto di una tabella modificandone le proprietà oppure applicando un TableLook. Ogni TableLook è costituito da un insieme di proprietà della tabella, ad esempio l'aspetto generale, le proprietà delle note a piè di pagina e delle celle e i bordi. È possibile selezionare uno dei TableLook predefiniti oppure creare e salvare un TableLook personalizzato.

TableLook

Un TableLook è un insieme di proprietà che definiscono l'aspetto di una tabella. È possibile selezionare un TableLook precedentemente definito oppure crearne uno personalizzato.

Prima o dopo aver applicato un TableLook, è possibile modificare i formati di singole celle o di gruppi di celle utilizzando le proprietà delle celle. I formati delle celle modificati verranno mantenuti anche se si applica un nuovo TableLook.

Ad esempio, è possibile iniziare applicando un TableLook *9POINT,* selezionare una colonna di dati e quindi, nella finestra di dialogo Formati di cella, impostare un carattere in grassetto per tale colonna. Successivamente, è possibile impostare il TableLook *BOXED*. La colonna precedentemente selezionata manterrà il carattere in grassetto, ma acquisirà le rimanenti caratteristiche del TableLook *BOXED*.

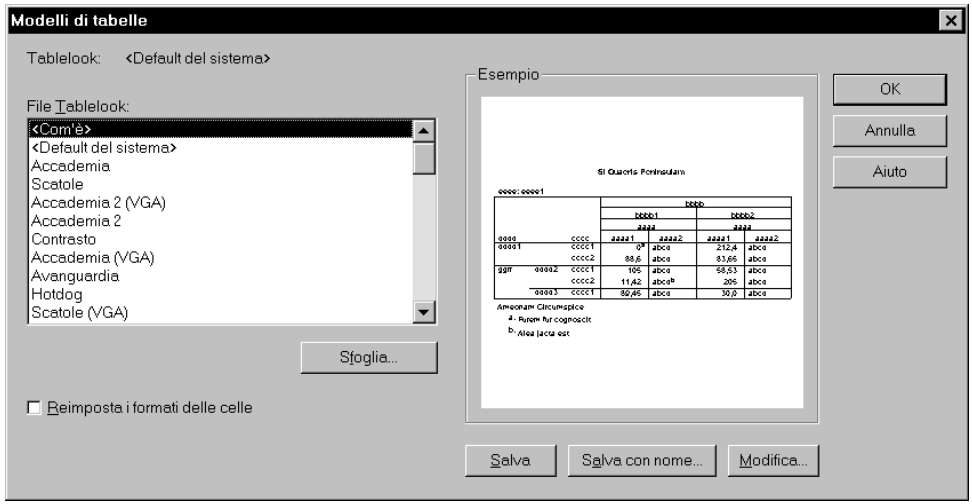
È inoltre possibile ripristinare in tutte le celle i formati definiti dal TableLook corrente. In questo modo verranno ripristinate tutte le celle modificate. Se nell'elenco di file TableLook è selezionata l'opzione Com'è, le celle modificate verranno reimpostate in base alle proprietà correnti della tabella.

Per applicare o salvare un TableLook

- Attivare una tabella pivot.
- Dai menu, scegliere:

  Formato
  Tablelook...

Figura 12-8
Finestra di dialogo TableLook



- Selezionare un TableLook dall'elenco dei file. Per selezionare un file in un'altra directory, fare clic su Sfoglia.
- Fare clic su OK per applicare il TableLook alla tabella pivot selezionata.

Per modificare o creare un TableLook

- Nella finestra di dialogo TableLooks selezionare un TableLook dall'elenco dei file.
- Fare clic su Modifica.
- Modificare le proprietà della tabella in base agli attributi desiderati e quindi fare clic su OK.
- Fare clic su Salva per salvare il TableLook modificato oppure su Salva con nome per salvarlo come nuovo TableLook.

  La modifica di un TableLook viene applicata solo alla tabella pivot selezionata. Un TableLook modificato verrà applicato alle tabelle per le quali è impostato solo se le tabelle vengono selezionate e vi viene applicato nuovamente il TableLook.

Proprietà tabella

Nella finestra di dialogo Proprietà tabella è possibile impostare le proprietà generali di una tabella, impostare gli stili delle celle per diverse parti della tabella e salvare un gruppo di proprietà come TableLook. Utilizzando le schede di questa finestra di dialogo è possibile:

- Impostare proprietà generali, ad esempio nascondere le righe o le colonne vuote e definire le proprietà di stampa.
- Impostare il formato e la posizione dei richiami alle note.
- Determinare formati specifici per le celle nell'area dei dati, per le etichette di riga e di colonna e per altre aree della tabella.
- Impostare lo spessore e il colore delle linee che formano i bordi di ciascuna area della tabella.

Per modificare le proprietà delle tabelle pivot

- ▶ Fare doppio clic su qualunque punto della tabella pivot per attivarla.
- ▶ Dai menu della tabella pivot, scegliere:
  Formato
  Proprietà tabella...
- ▶ Selezionare una scheda (Generale, Note a piè di pagina, Formati di cella, Bordi o Stampa).
- ▶ Selezionare le opzioni desiderate.
- ▶ Fare clic su OK o su Applica.

Le nuove proprietà vengono applicate alla tabella pivot selezionata. Per applicare le nuove proprietà della tabella a un TableLook anzichè alla tabella selezionata, modificare il TableLook (menu Formato, TableLook).

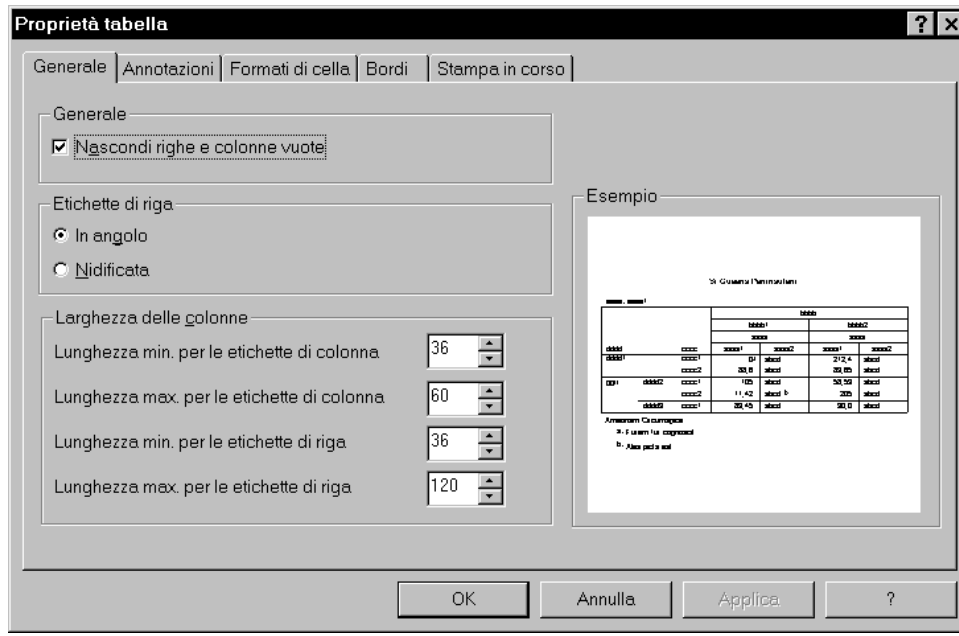
Proprietà tabella: Generale

Diverse proprietà vengono applicate all'intera tabella. È possibile:

- Visualizzare o nascondere le righe e le colonne vuote. Una riga o una colonna vuota contiene celle di dati vuote.

- Controllare la posizione delle etichette riga, che possono essere nell'angolo in alto a sinistra o nidificate.
- Impostare la larghezza minima e massima della colonna (espressa in punti).

Figura 12-9
Finestra di dialogo Proprietà tabella, scheda Generale



Per modificare le proprietà generali di una tabella

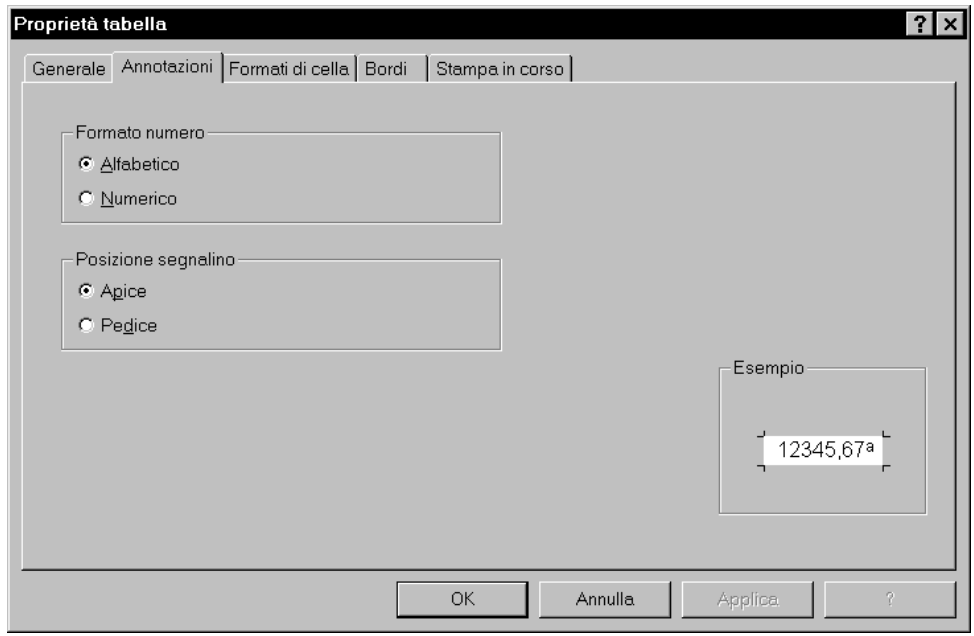
- ▶ Fare clic sulla scheda Generale.
- ▶ Selezionare le opzioni desiderate.
- ▶ Fare clic su OK o su Applica.

Proprietà tabella: Note a piè di pagina

Le proprietà dei rimandi alle note includono lo stile e la posizione in relazione al testo.

- Lo stile dei rimandi alle note può essere rappresentato da numeri (1, 2, 3...) o da lettere (a, b, c...).
- I rimandi alle note possono essere associati al testo in formato apice o pedice.

Figura 12-10
Finestra di dialogo Proprietà tabella, scheda Note a piè di pagina



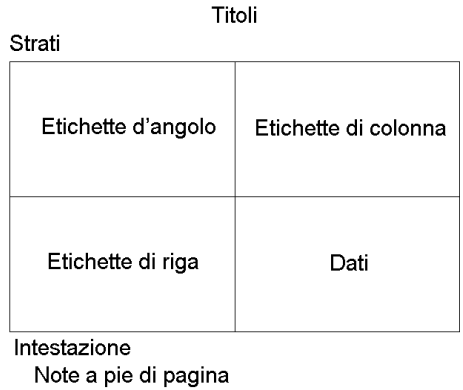
Per modificare le proprietà dei rimandi alle note

- ▶ Selezionare la scheda Note a piè di pagina.
- ▶ Selezionare un formato per i rimandi alle note.
- ▶ Selezionare la posizione dei rimandi.
- ▶ Fare clic su OK o su Applica.

Proprietà tabella: Formati di cella

Per quanto riguarda la formattazione, la tabella è suddivisa in più aree: Titolo, Strati, Etichette angoli, Etichette righe, Etichette colonne, Dati, Didascalia e Note a piè di pagina. In ogni area della tabella è possibile modificare i formati di cella corrispondenti. I formati di cella includono caratteristiche (carattere, dimensione, colore, stile) del testo, allineamento orizzontale e verticale, ombreggiatura della cella, colori di primo piano e di sfondo e margini interni della cella.

Figura 12-11
Aree di una tabella

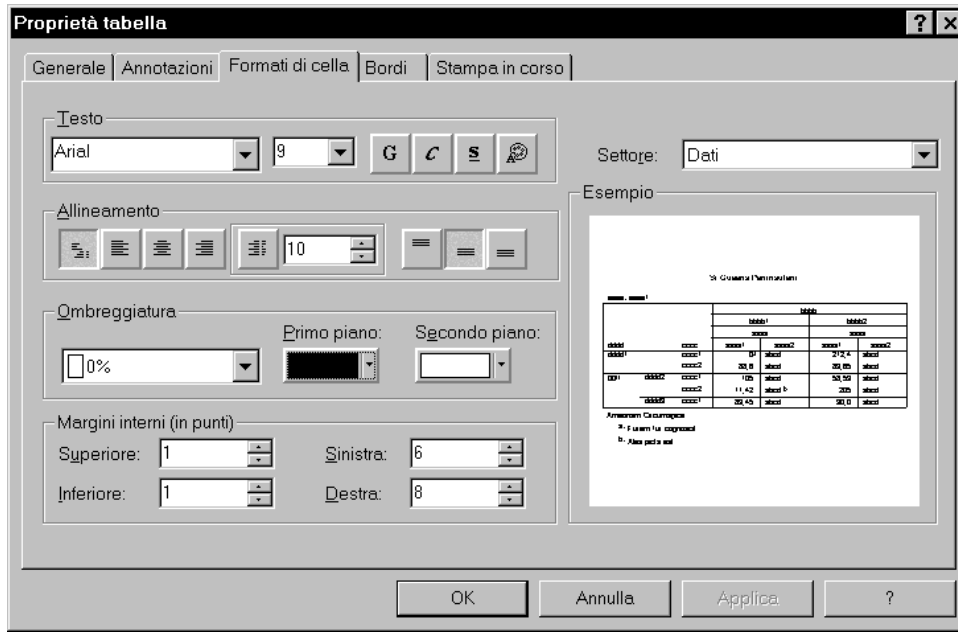


I formati di cella vengono applicati alle aree (categorie di informazioni). I formati non sono caratteristici di singole celle. Questa distinzione è importante ai fini del pivoting di una tabella.

Ad esempio:

- Se si specifica un carattere in grassetto come formato di cella delle etichette di colonna, queste appariranno in grassetto indipendentemente dalle informazioni visualizzate nella dimensione colonna. Se l'elemento viene spostato dalla dimensione colonna a una diversa dimensione non manterrà la caratteristica grassetto delle etichette di colonna.
- Se alle etichette delle colonne viene applicato il grassetto semplicemente evidenziando le celle di una tabella pivot attivata e facendo clic sul pulsante Grassetto sulla barra degli strumenti, il contenuto delle celle rimarrà in grassetto indipendentemente dalla dimensione in cui viene spostato, mentre nelle etichette delle colonne la caratteristica grassetto non verrà applicata ad altri elementi spostati nella dimensione colonna.

Figura 12-12
Finestra di dialogo Proprietà tabella, scheda Formato celle



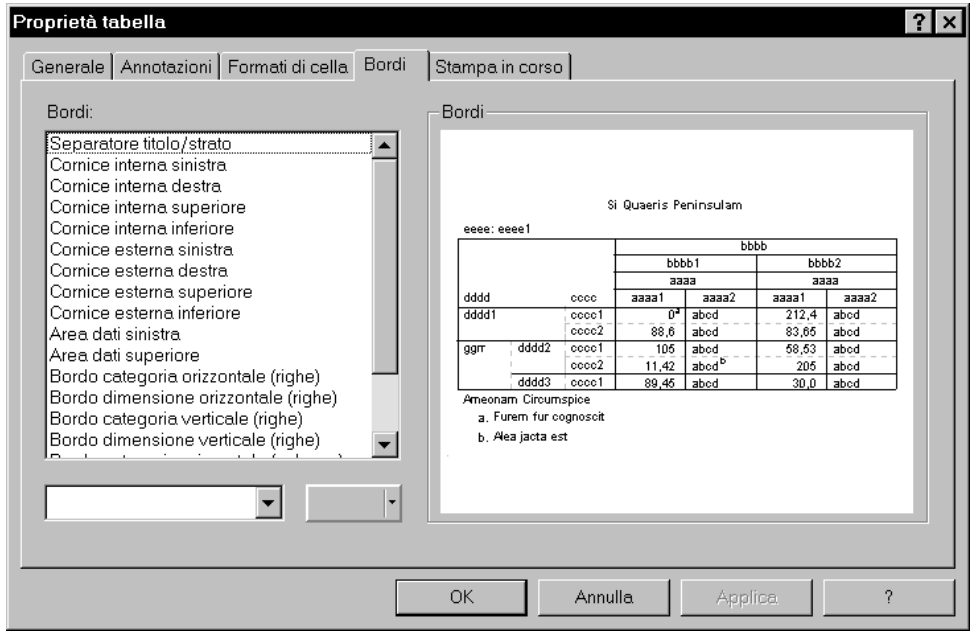
Per modificare i formati di cella

- Selezionare la scheda Formati di cella.
- Selezionare un'area dall'elenco a discesa oppure fare clic su un area del campione.
- Selezionare le caratteristiche dell'area. Le selezioni verranno applicate al campione.
- Fare clic su OK o su Applica.

Proprietà tabella: Bordi

È possibile selezionare uno stile di linea e un colore da applicare nella posizione di ciascun bordo della tabella. Se come stile si seleziona Nessuno, nella posizione selezionata non verrà inserita alcuna linea.

Figura 12-13
Finestra di dialogo Proprietà tabella, scheda Bordi



Per modificare i bordi di una tabella

- Fare clic sulla scheda Bordi.
- Selezionare la posizione del bordo facendo clic sul suo nome nell'elenco oppure facendo clic su una linea nell'area Esempio. Fare clic tenendo premuto Maiusc per selezionare più nomi oppure fare clic tenendo premuto Ctrl per selezionare nomi non contigui.
- Selezionare uno stile di linea oppure Nessuno.
- Selezionare un colore.
- Fare clic su OK o su Applica.

Per visualizzare i bordi nascosti in una tabella pivot

Per le tabelle in cui non sono visibili molti bordi, è possibile visualizzare i bordi nascosti. In questo modo è possibile eseguire più rapidamente alcune operazioni, ad esempio la modifica della larghezza delle colonne. I bordi nascosti (griglia) vengono visualizzati nel Viewer ma non vengono stampati.

- ▶ Fare doppio clic su qualunque punto della tabella pivot per attivarla.
- ▶ Dai menu, scegliere:

  Visualizza
  Griglia

Proprietà tabella: Stampa

Per le tabelle pivot stampate è possibile definire quanto segue:

- Stampare tutti gli strati o solo lo strato superiore di una tabella e stampare ciascuno strato in una pagina distinta. Questa impostazione influisce solo sulla stampa e non sulla visualizzazione degli strati nel Viewer.
- Ridurre una tabella orizzontalmente o verticalmente in base alle dimensioni della pagina stampata.
- Controllare le righe vedove/orfane, ovvero il numero minimo di righe e colonne che verranno incluse in una sezione stampata di una tabella, se questa ha lunghezza o larghezza eccessive per le dimensioni definite per la pagina. (*Nota*: Se una tabella è troppo lunga per poter essere inclusa nella pagina corrente solo perché è preceduta da altro output, verrà stampata automaticamente in una nuova pagina, indipendentemente dall'impostazione delle righe vedove/orfane.
- Includere testo di prosecuzione per le tabelle che non rientrano in una sola pagina. È possibile visualizzare il testo di prosecuzione in testa o in coda a ciascuna pagina. Se nessuna delle due opzioni è selezionata, il testo di prosecuzione non sarà visualizzato.

Per impostare la stampa di tabelle pivot

- ▶ Fare clic sulla scheda Stampa.
- ▶ Selezionare le opzioni di stampa desiderate.

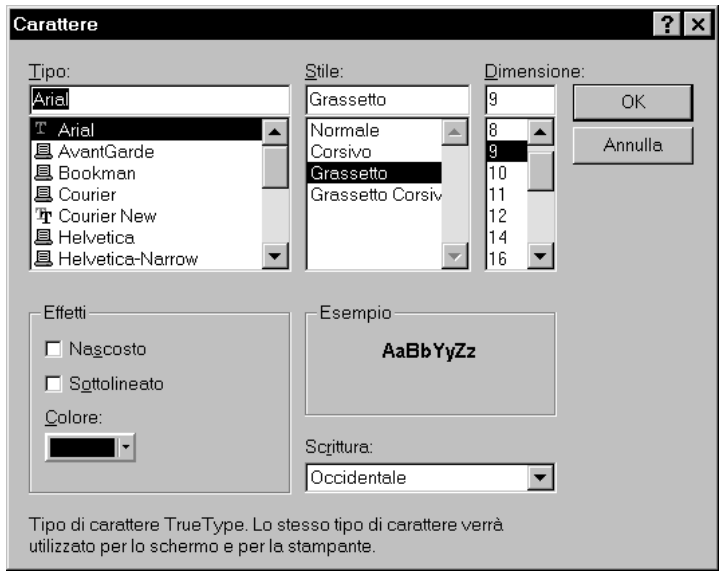
- Fare clic su OK o su Applica.

Carattere

Utilizzando un TableLook è possibile specificare gli attributi dei caratteri in diverse aree della tabella. È inoltre possibile modificare il tipo di carattere nelle singole celle. Le opzioni per il carattere nelle celle includono tipo di carattere, stile, dimensione e colore. È inoltre possibile nascondere il testo o sottolinearlo.

Se si specificano le proprietà del carattere di una cella, queste verranno applicate a tutti gli strati della tabella che includono la stessa cella.

Figura 12-14
Finestra di dialogo Carattere



Per modificare il carattere in una cella

- Attivare la tabella pivot e selezionare il testo che si desidera modificare.
- Dai menu della tabella pivot, scegliere:

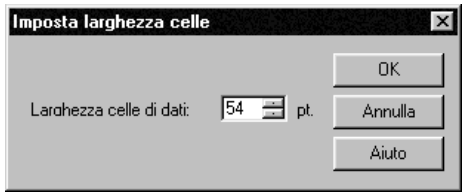
  Formato
  Carattere...

In alternativa è possibile anche selezionare un carattere, uno stile di carattere nonché scegliere un colore e uno stile per lo script e decidere se nascondere o sottolineare il testo.

Larghezza delle celle

L'opzione Imposta larghezza celle consente di impostare la stessa larghezza per tutte le celle di dati.

Figura 12-15
Finestra di dialogo Imposta larghezza celle



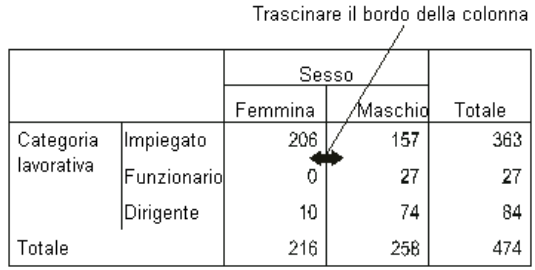
Per modificare la larghezza delle celle di dati

- Attivare la tabella pivot.
- Dai menu, scegliere:
  Formato
  Imposta larghezza celle...
- Specificare il valore della larghezza delle celle.

Per modificare la larghezza di una colonna della tabella pivot

- Fare doppio clic sulla tabella per attivarla.
- Spostare il puntatore del mouse sulle etichette di categoria fino a posizionarlo sul bordo destro della colonna che si desidera modificare. Il puntatore assumerà la forma di una freccia a due punte.
- Tenere premuto il pulsante del mouse mentre si trascina il bordo nella posizione desiderata.

Figura 12-16
Modifica della larghezza di una colonna



È possibile modificare i bordi verticali delle categorie e delle dimensioni nell'area delle etichette delle righe, indipendentemente dal fatto che siano visualizzati.

- Spostare il puntatore del mouse sulle etichette di riga finché non assume la forma di una freccia a due punte.
- Trascinare il bordo fino ad ottenere la larghezza desiderata.

Proprietà cella

Le opzioni di Proprietà cella vengono applicate a una singola cella selezionata. È possibile modificare il formato dei valori, l'allineamento, i margini e l'ombreggiatura. Le opzioni di Proprietà cella hanno la priorità sulle opzioni di Proprietà tabella. Pertanto, se si modificano le proprietà della tabella, non vengono modificate le proprietà delle celle applicate singolarmente.

Per modificare le proprietà delle celle

- Attivare una tabella e selezionare una cella al suo interno.
- Dai menu, scegliere:

Formato
  Proprietà cella...

Proprietà cella: Valore

In questa scheda della finestra di dialogo è possibile impostare il formato dei valori di una cella. È possibile selezionare il formato dei numeri, della data, dell'orario e della valuta e impostare il numero di cifre decimali da visualizzare.

Figura 12-17
Finestra di dialogo Proprietà celle, scheda Valore

Per modificare i formati di valore di una cella

- Fare clic sulla scheda Valore.
- Selezionare una categoria e un formato.
- Selezionare il numero di cifre decimali.

Per modificare i formati dei valori di una colonna

- Fare clic tenendo premuti Ctrl e Alt sull'etichetta della colonna.

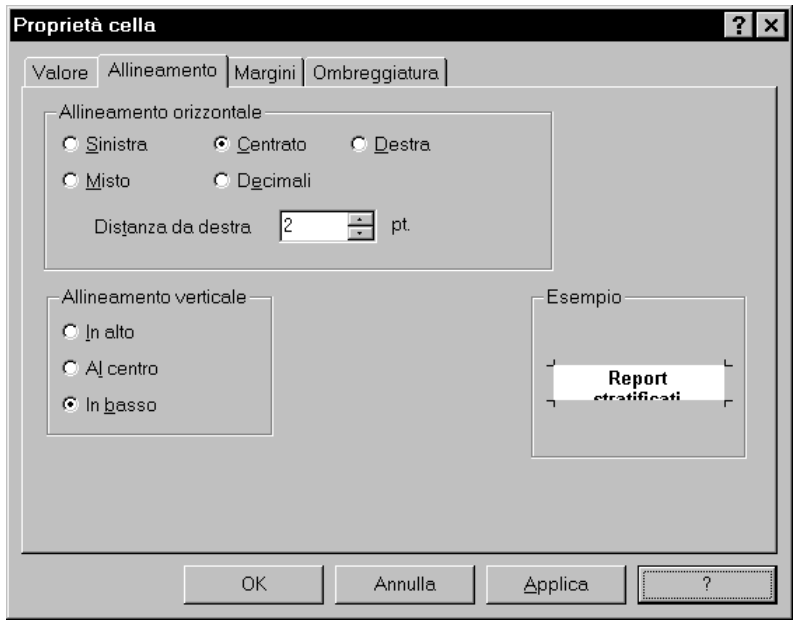
- Fare clic con il pulsante destro del mouse sulla colonna evidenziata.
- Dal menu di scelta rapida selezionare Proprietà cella.
- Fare clic sulla scheda Valore.
- Selezionare il formato che si desidera applicare alla colonna.

È possibile utilizzare questo metodo per eliminare o aggiungere segni di percentuale e segni di dollaro, modificare il numero di decimali visualizzati e passare dalla notazione scientifica alla normale visualizzazione numerica.

Proprietà cella: Allineamento

In questa scheda della finestra di dialogo è possibile impostare l'allineamento orizzontale e verticale e la direzione del testo per una cella. Se si sceglie Misto, il contenuto della cella verrà allineato in base al tipo (numero, data o testo).

Figura 12-18
Finestra di dialogo Proprietà celle, scheda Allineamento



Per modificare l'allineamento nelle celle

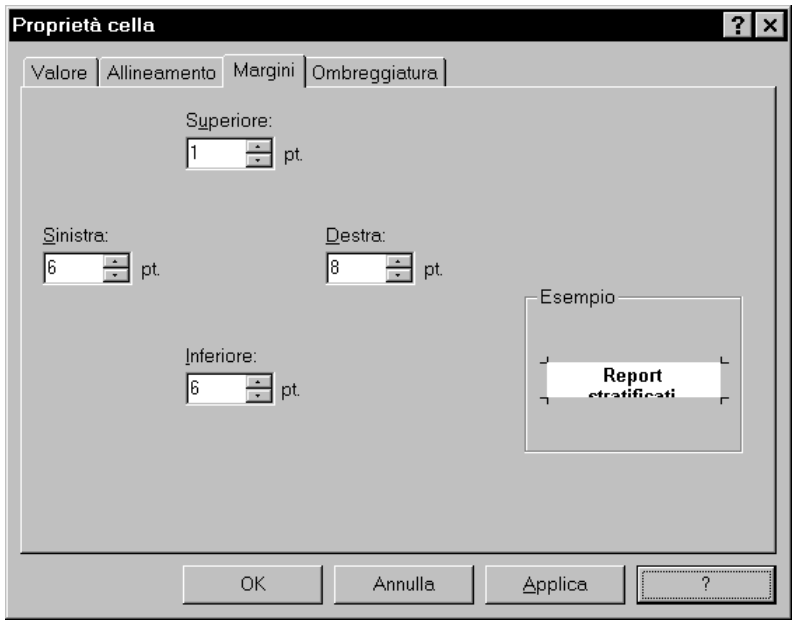
- Selezionare una cella nella tabella.
- Dai menu della tabella pivot, scegliere:
  Formato
  Proprietà cella...
- Fare clic sulla scheda Allineamento.

  Quando si selezionano le proprietà di allineamento per la cella, queste vengono illustrate nell'area Esempio.

Proprietà cella: Margini

In questa finestra di dialogo è possibile specificare la posizione di ciascun margine della cella.

Figura 12-19
Finestra di dialogo Proprietà celle, scheda Margini



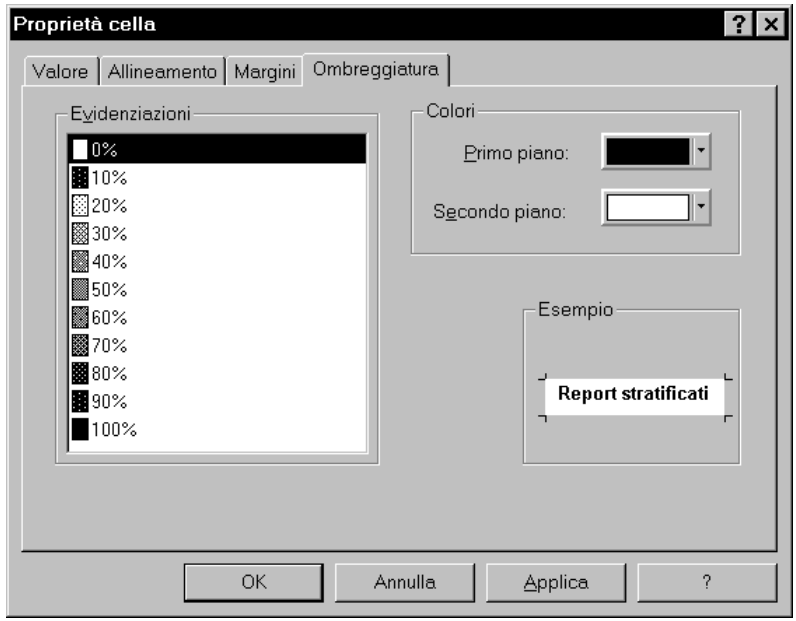
Per modificare i margini di una cella

- Fare clic sulla scheda Margini.
- Selezionare la posizione di ciascuno dei quattro margini.

Proprietà cella: Ombreggiatura

In questa finestra di dialogo è possibile specificare la percentuale di ombreggiatura e i colori di primo piano e di sfondo per l'area della cella selezionata. Questa impostazione non modifrica il colore del testo.

Figura 12-20
Finestra di dialogo Proprietà celle, scheda Ombreggiatura



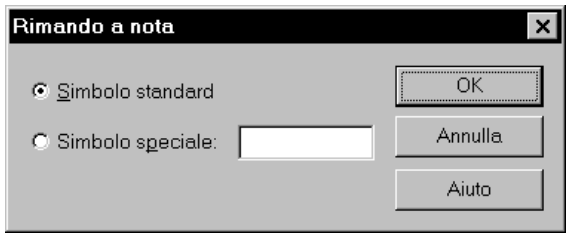
Per modificare l'ombreggiatura delle celle

- Fare clic sulla scheda Ombreggiatura.
- Selezionare le evidenziazioni e i colori che si desidera applicare alla cella.

Rimando a nota

L'opzione Rimando a nota consente di modificare i caratteri utilizzati nei rimandi alle note.

Figura 12-21
Finestra di dialogo Rimando a nota



Per modificare i caratteri dei rimandi alle note

- Selezionare una nota a piè di pagina.
- Dai menu della tabella pivot, scegliere:
  Formato
  Rimando a nota...
- Digitare uno o due caratteri.

Per rinumerare le note a piè di pagina

Dopo aver effettuato il pivoting di una tabella modificando righe, colonne e strati, le note a piè di pagina possono risultare ordinate diversamente. Per rinumerare le note a piè di pagina:

- Attivare la tabella pivot.
- Dai menu, scegliere:
  Formato
  Rinumera note a piè di pagina

Selezione di righe e colonne nelle tabelle pivot

Le tabelle pivot pongono alcune limitazioni sulla selezione di intere righe e colonne e l'evidenziazione che indica la riga o la colonna selezionata può essere estesa ad aree non contigue della tabella.

Per selezionare una riga o una colonna in una tabella pivot

- Fare doppio clic su qualunque punto della tabella pivot per attivarla.
- Fare clic su un'etichetta di riga o di colonna.
- Dai menu, scegliere:

  Modifica
  Selezionare
  Celle dati e celle etichette

  oppure

- Tenere premuti Ctrl e Alt e fare clic sull'etichetta di riga o di colonna.

Se la tabella contiene più di una dimensione nell'area della riga o della colonna, la sezione evidenziata potrebbe estendersi su più celle non contigue.

Modifica dei risultati della tabella pivot

Nel Viewer il testo viene visualizzato in più elementi. È possibile aggiungere nuovo testo oppure modificare quello esistente. Per modificare le tabelle pivot è possibile:

- Modificare il testo nelle celle della tabella pivot.
- Aggiungere didascalie e note a piè di pagina.

Per modificare il testo in una cella

- Attivare la tabella pivot.
- Fare doppio clic sulla cella.
- Modificare il testo.

▶ Premere Invio per registrare le modifiche oppure premere Esc per ripristinare il contenuto precedente della cella.

Per aggiungere didascalie a una tabella

▶ Dai menu della tabella pivot, scegliere:

Inserisci
Didascalia

Alla fine della tabella viene visualizzata la dicitura Didascalia tabella.

▶ Fare doppio clic sulle parole Didascalia tabella e sostituirle con il testo desiderato.

Per aggiungere una nota a piè di pagina a una tabella

È possibile aggiungere una nota a piè di pagina a qualsiasi elemento di una tabella.

▶ Fare clic su un titolo, su una cella o su una didascalia all'interno di una tabella pivot attivata.

▶ Dai menu della tabella pivot, scegliere:

Inserisci
Nota a piè di pagina...

▶ Fare doppio clic sulla parola Nota a piè di pagina e sostituirla con il testo desiderato.

Stampa delle tabelle pivot

I fattori che influenzano la stampa delle tabelle pivot sono molteplici e possono essere definiti modificando gli attributi delle tabelle pivot.

- Per le tabelle pivot multidimensionali (tabelle con strati), è possibile scegliere se stampare tutti gli strati o soltanto il primo strato (visibile).
- Per le tabelle pivot lunghe o grandi, è possibile ridefinirne automaticamente le dimensioni in modo da adattarle alla pagina oppure impostare con esattezza le interruzioni delle tabelle o delle pagine.

Scegliere Anteprima di stampa del menu File per visualizzare le tabelle pivot prima di stamparle.

Per stampare gli strati nascosti di una tabella pivot

- ▶ Fare doppio clic sulla tabella per attivarla.
- ▶ Dai menu, scegliere:
  Formato
  Proprietà tabella...
- ▶ Nella scheda Stampa selezionare Stampa tutti gli strati.

  È inoltre possibile stampare ciascuno strato della tabella pivot in una pagina distinta.

Impostazione delle interruzioni di pagina per le tabelle con larghezza o lunghezza eccessiva

Le tabelle pivot con larghezza o altezza eccessiva per essere stampate entro le dimensioni definite per la pagina verranno divise automaticamente e stampate in più sezioni. Per le tabelle di larghezza eccessiva, le diverse sezioni verranno stampate nella stessa pagina se è disponibile spazio sufficiente. È possibile:

- Impostare i punti delle righe e delle colonne in cui le tabelle vengono divise.
- Specificare le righe e le colonne che devono essere mantenute insieme quando le tabelle vengono divise.
- Ridimensionare le tabelle di grandi dimensioni in modo da adattarle alla pagina.

Per specificare le interruzioni di riga e di colonna per le tabelle pivot da stampare

- ▶ Attivare la tabella pivot.
- ▶ Fare clic sull'etichetta di colonna o di riga situata rispettivamente a sinistra o al di sopra del punto in cui si desidera inserire l'interruzione.
- ▶ Dai menu, scegliere:
  Formato
  Inserisci Interrompi qui

Per specificare le righe o le colonne da mantenere unite

- Attivare la tabella pivot.
- Selezionare le etichette delle righe o delle colonne che si desidera mantenere unite. Fare clic e trascinare oppure fare clic tenendo premuto Maiusc per selezionare più etichette di riga o di colonna.
- Dai menu, scegliere:

  Formato
  Mantieni unito

Per ridimensionare una tabella adattandola alla pagina

- Attivare la tabella pivot.
- Dai menu, scegliere:

  Formato
  Proprietà tabella
- Fare clic sulla scheda Stampa.
- Fare clic su Ridimensiona la tabella in larghezza.

  e/o
- Fare clic su Ridimensiona la tabella in lunghezza.

Capitolo
13
Utilizzo della sintassi dei comandi

ghhggh Inoltre, offre alcune funzionalità a cui non è possibile accedere tramite i menu e le finestre di dialogo.

È possibile accedere alla maggior parte dei comandi tramite i menu e le finestre di dialogo. Tuttavia, alcuni comandi e opzioni sono disponibili solo utilizzando il linguaggio dei comandi. Questo linguaggio consente inoltre di salvare le sessioni in un file di sintassi, in modo da poter ripetere l'analisi successivamente oppure eseguirla in una sessione automatizzata con l'utilità SPSS Production.

Un file di sintassi è un semplice file di testo che contiene comandi. Sebbene sia possibile aprire una finestra di sintassi e digitarvi dei comandi, risulta senz'altro più semplice lasciare al programma il compito di creare il file di sintassi utilizzando uno dei seguenti metodi:

- Incollando la sintassi dei comandi dalle finestre di dialogo
- Copiando la sintassi dal registro dell'output
- Copiando la sintassi dal file giornale

Informazioni dettagliate sui riferimenti alla sintassi dei comandi sono disponibili in due formati: nella Guida integrata e su un documento PDF a parte in *SPSS Command Syntax Reference*, anch'esso consultabile tramite il menu ?.

Per visualizzare la Guida sensibile al contesto per il comando visualizzato nella finestra della sintassi, premere il tasto F1.

Regole della sintassi dei comandi

Quando si eseguono i comandi dalla finestra della sintassi di comando durante una sessione SPSS, i comandi vengono eseguiti in modalità **interattiva**.

Le seguenti regole si applicano alle specifiche dei comandi in modalità interattiva:

- Ciascun comando deve terminare con un punto. È sempre consigliabile omettere la fine del comando con `BEGIN DATA` per essere certi che i dati vengono considerati come specifica continua.
- La fine del comando deve sempre essere l'ultimo carattere del comando, spazi esclusi.
- Se il comando non termina con un punto, la linea vuota viene interpretata come fine del comando.
- I comandi possono iniziare in qualsiasi colonna di una riga di comando e continuare per il numero di righe necessario. L'unica eccezione è rappresentata dal comando `END DATA`, che deve iniziare nella prima colonna della prima riga dopo la fine dei dati.

Nota: per garantire la compatibilità con le altre modalità di esecuzione dei comandi (compresi i file di comando eseguiti con i comandi `INSERT` o `INCLUDE` in modalità interattiva, è indispensabile che ciascuna riga della sintassi di comando non superi 256 byte.

- La maggior parte dei sottocomandi vengono separati da barre (/). La barra che precede il primo sottocomando di un comando è in genere opzionale.
- I nomi delle variabili devono essere riportati per intero.
- Il testo racchiuso tra apostrofi o virgolette deve occupare una sola riga.
- Indipendentemente dalle impostazioni internazionali di Windows, il separatore dei decimali deve essere un punto (.).
- I nomi di variabili che terminano con un punto possono causare errori nei comandi creati nelle finestre di dialogo. Nelle finestre di dialogo non è possibile creare nomi di variabili con un punto finale ed in genere è necessario evitarne l'utilizzo.

Per la sintassi dei comandi non si fa distinzione tra maiuscole e minuscole e, per molti dei comandi, è possibile utilizzare delle abbreviazioni di tre o quattro lettere. Per specificare un singolo comando è possibile utilizzare un numero illimitato di righe. È possibile aggiungere uno spazio o delle interruzioni di riga in qualsiasi punto in cui sia ammesso l'utilizzo di un singolo spazio vuoto, come prima e dopo le barre, le parentesi e gli operatori aritmetici o tra i nomi di variabili. Ad esempio,

```
FREQUENCIES
  VARIABLES=JOBCAT GENDER
```

```
  /PERCENTILES=25 50 75
  /BARCHART.
```

che

```
freq var=jobcat gender /percent=25 50 75 /bar.
```

sono entrambe alternative accettabili che generano i medesimi risultati.

File INCLUDE

Per i comandi eseguiti tramite il comando `INCLUDE` si applicano le regole della sintassi della modalità batch.

Le seguenti regole si applicano a specifiche di comandi in modalità batch o produzione:

- tutti i comandi nel file di comandi devono iniziare alla colonna 1; è possibile usare i simboli più (+) o meno (-) nella prima colonna per inserire un rientro per la specifica del comando in modo da renderlo più leggibile.
- Se si utilizzano più righe per il comando, la colonna 1 di ciascuna riga successiva deve essere lasciata vuota.
- I terminatori dei comandi sono opzionali.
- Ciascuna rigadeve avere una lunghezza massima di 256 byte; i caratteri aggiuntivi vengono tagliati.

Se non si hanno file di comandi che già utilizzano il comando`INCLUDE`, è preferibile usare il comando `INSERT` poiché può essere utilizzato con file di comando conformi ad entrambi gli insiemi di regole. Se la sintassi viene generata incollando le selezioni nelle finestre di dialogo in una finestra di sintassi, il formato dei comandi è idoneo per qualsiasi modalità. Per ulteriori informazioni, vedere *SPSS Command Syntax Reference*, disponibili in formato PDF nel menu ?.

Creazione di sintassi dei comandi nelle finestre di dialogo

Il metodo più semplice per creare un file di sintassi consiste nell'eseguire selezioni nelle finestre di dialogo di SPSS e quindi incollare la sintassi per le selezioni in una finestra di sintassi. Incollando la sintassi in ciascuna fase di una lunga analisi, si

costruisce un file della sessione che permette di ripetere l'analisi in un secondo momento o di eseguire una sessione automatizzata con l'utilità SPSS Production.

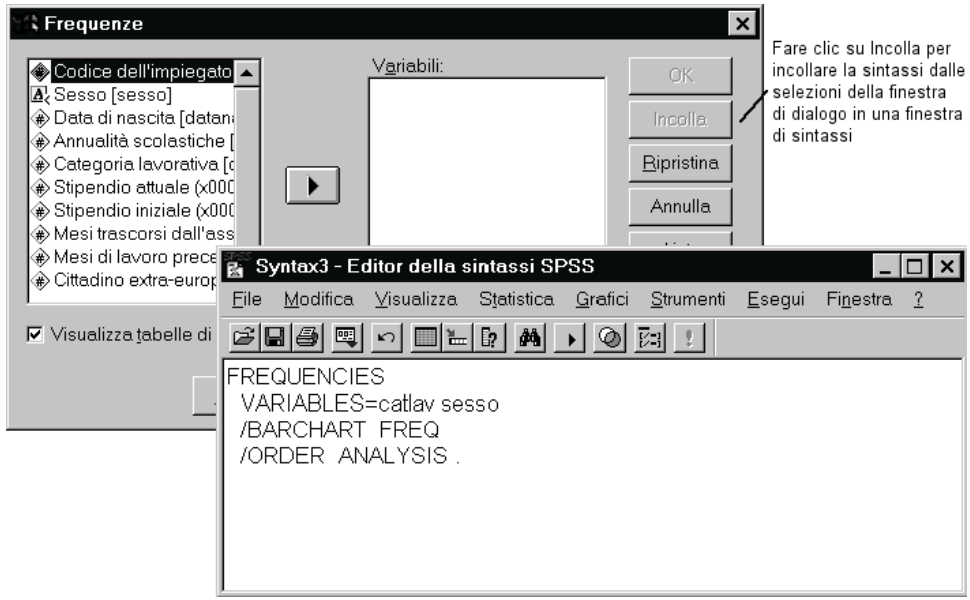
Nella finestra di sintassi, è possibile eseguire la sintassi incollata, modificarla e salvarla in un file di sintassi.

Per creare sintassi dei comandi nelle finestre di dialogo

- Aprire la finestra di dialogo ed eseguire tutte le selezioni desiderate.
- Fare clic su Incolla.

La sintassi dei comandi viene incollata nella finestra di sintassi designata. Se non è aperta alcuna finestra di sintassi, ne viene automaticamente aperta una nuova in cui viene incollata la sintassi copiata.

Figura 13-1
Sintassi incollata da una finestra di dialogo



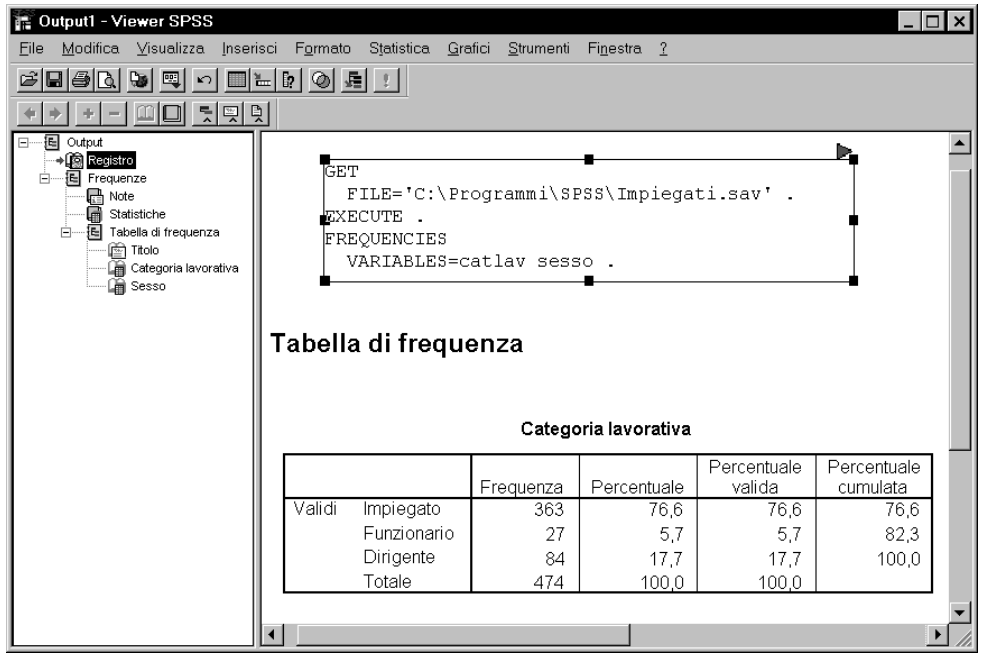
Nota: Quando la finestra di dialogo viene aperta tramite i menu di una finestra di script, il codice per l'esecuzione della sintassi da uno script viene incollato nella finestra di script.

Copia della sintassi dal registro dell'output

È possibile costruire un file di sintassi copiando la sintassi dei comandi dal registro visualizzato nel Viewer. Per utilizzare questo metodo, selezionare Visualizza i comandi nel registro nelle impostazioni del Viewer (menu Modifica, Opzioni, scheda Viewer) prima di eseguire l'analisi. In tal modo, tutti i comandi appariranno nel Viewer insieme all'output dell'analisi.

Nella finestra di sintassi, è possibile eseguire la sintassi incollata, modificarla e salvarla in un file di sintassi.

Figura 13-2
Sintassi dei comandi nel registro



Per copiare la sintassi dal registro dell'output

- Prima di eseguire l'analisi, dai menu scegliere:

Modifica
Opzioni...

- Nella scheda Viewer, selezionare Visualizza i comandi nel registro.

  All'esecuzione delle analisi, i comandi relativi alle selezioni effettuate nella finestra di dialogo vengono inseriti nel registro.

- Aprire un file di sintassi salvato in precedenza oppure crearne uno nuovo. Per creare un nuovo file di sintassi, dai menu scegliere:

  File
  Nuovo
  Sintassi

- Nel Viewer, fare doppio clic su una voce del registro per attivarla.

- Fare clic e trascinare il mouse in modo da evidenziare la sintassi da copiare.

- Dai menu del Viewer scegliere:

  Modifica
  Copia

- In una finestra di sintassi, dai menu scegliere:

  Modifica
  Incolla

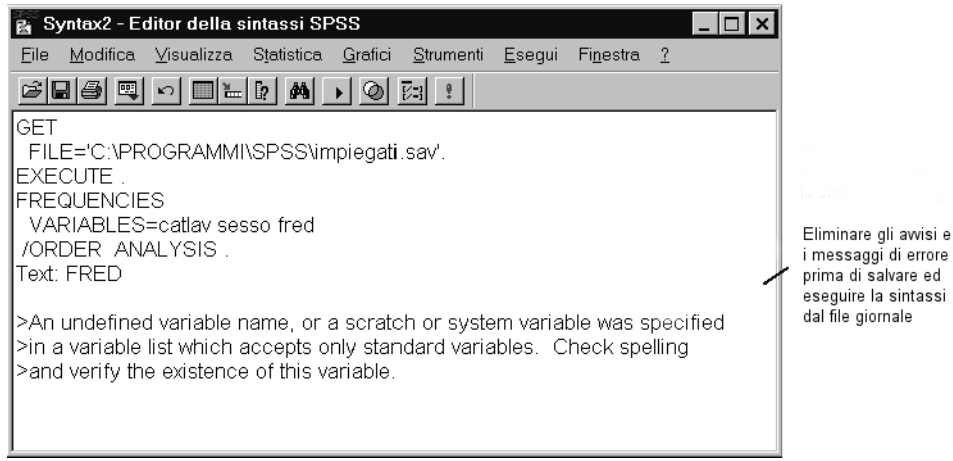
Utilizzo di sintassi di un file giornale

Per impostazione predefinita, tutti i comandi eseguiti durante una sessione vengono registrati in un file giornale denominato *spss.jnl* (impostato tramite il comando Opzioni del menu Modifica). Questo file può essere modificato e salvato come file di sintassi da utilizzare per ripetere un'analisi già eseguita o per eseguirla in una sessione automatizzata con l'utilità SPSS Production.

Il file giornale è un file di testo modificabile come un qualsiasi altro file di testo. Poiché nel file giornale, oltre alle sintassi dei comandi, vengono registrati anche messaggi di errore e avvisi, è necessario eliminare tali messaggi prima di salvare il file di sintassi. Per una corretta esecuzione della sessione, è ovviamente necessario correggere gli eventuali errori.

Memorizzare il file giornale modificato con un altro nome. Il file giornale viene automaticamente esteso o sovrascritto a ogni sessione e quindi, se si utilizzasse lo stesso nome per il file di sintassi, si potrebbero ottenere dei risultati imprevisti.

Figura 13-3
Modifica del file giornale


Syntax2 - Editor della sintassi SPSS
File Modifica Visualizza Statistica Grafici Strumenti Esegui Finestra ?
GET
FILE='C:\PROGRAMMI\SPSS\impiegati.sav'.
EXECUTE .
FREQUENCIES
VARIABLES=catlav sesso fred
/ORDER ANALYSIS .
Text: FRED
>An undefined variable name, or a scratch or system variable was specified
>in a variable list which accepts only standard variables. Check spelling
>and verify the existence of this variable.
Eliminare gli avvisi e i messaggi di errore prima di salvare ed eseguire la sintassi dal file giornale

Per utilizzare la sintassi di un file giornale

- Per aprire il file giornale, dai menu scegliere:

  File
  Apri
  Altro...

- Individuare e aprire il file giornale (per impostazione predefinita, il file *spss.jnl* si trova nella directory *temp*).

  Selezionare Tutti i file (*.*) in Tipo file oppure immettere *.jnl nella casella Nome file in modo da visualizzare i file giornale nell'elenco dei file. Se non si riesce a trovare il file, utilizzare il comando Opzioni del menu Modifica per verificare che il giornale venga memorizzato all'interno del sistema.

- Modificare il file eliminando gli eventuali messaggi di errore o avvisi, indicati dal simbolo >.

- Salvare il file giornale modificato con un altro nome. È consigliabile utilizzare un nome di file con l'estensione *sps*, quella predefinita per i file giornale.

Per eseguire la sintassi dei comandi

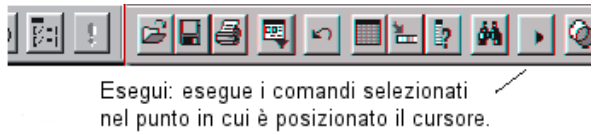
- Evidenziare i comandi che si desidera eseguire nella finestra della sintassi.
- Fare clic sul pulsante Esegui (il triangolo rivolto verso destra) nella barra degli strumenti dell'Editor della sintassi.

  oppure

- Selezionare una delle opzioni del menu Esegui.
  - **Tutto.** Vengono eseguiti tutti i comandi presenti nella finestra di sintassi.
  - **Selezione.** Vengono eseguiti solo i comandi correntemente selezionati, compresi i comandi parzialmente evidenziati.
  - **Corrente.** Viene eseguito solo il comando su cui è attualmente posizionato il cursore.
  - **Fino alla fine.** Vengono eseguiti tutti i comandi a partire dalla posizione corrente del cursore e fino alla fine del file di sintassi dei comandi.

  Il pulsante Esegui della barra degli strumenti dell'Editor della sintassi consente di eseguire i comandi selezionati oppure, se non vi è alcun comando selezionato, il comando su cui è posizionato il cursore.

Figura 13-4
Barra degli strumenti dell'Editor della sintassi



Comandi EXECUTE multipli

La sintassi copiata dalle finestre di dialogo o copiata dal registro o dal giornale potrebbe contenere i comandi `EXECUTE`. Se si eseguono più comandi da una finestra di sintassi, non sono necessari più comandi `EXECUTE` che potrebbero ridurre le prestazioni perché tale comando legge tutto il file di dati.

- Se l'ultimo comando nel file di sintassi è un comando che legge il file di dati (come una procedura grafica o statistica), nessun comando EXECUTE sarà necessario e potrà essere eliminato.
- Se non si è certi che l'ultimo comando legga il file di dati, nella maggior parte dei casi è possibile eliminare tutto eccetto l'ultimo comando EXECUTE nel file di sintassi.

Ritardi

Un'eccezione interessante è rappresentata dai comandi di trasformazione che contengono i ritardi. In una serie di comandi di trasformazione senza l'impiego dei comandi EXECUTE o di altri comandi per la lettura dei dati, i ritardi vengono calcolati dopo tutte le altre trasformazioni, indipendentemente dall'ordine dei comandi. Ad esempio:

```
COMPUTE lagvar=LAG(var1)
COMPUTE var1=var1*2
```

che

```
COMPUTE lagvar=LAG(var1)
EXECUTE
COMPUTE var1=var1*2
```

Generano risultati molto diversi per il valore di *lagvar*, poiché il primo comando utilizza il valore trasformato di *var1* mentre il secondo usa il valore originale.

Capitolo
14

Frequenze

La procedura Frequenze consente di ottenere statistiche e rappresentazioni grafiche che risultano utili per la descrizione di molti tipi di variabili. La procedura Frequenza offre un'ottima opportunità per iniziare ad osservare i dati.

Per ottenere un rapporto e un grafico a barre delle frequenze è possibile disporre i singoli valori in ordine crescente o decrescente oppure ordinare le categorie in base alle rispettive frequenze. Il rapporto sulle frequenze può essere eliminato se una variabile ha molti valori distinti. È possibile etichettare i grafici con frequenze (default) o percentuali.

Esempio. Qual è la distribuzione dei clienti di un'azienda per tipo di industria? Dall'output, si nota che il 37,5% dei clienti fa parte di enti governativi, il 24,9% fa parte di società, il 28,1% di istituzioni accademiche e il 9,4% del settore sanitario. Per i dati quantitativi e continui, ad esempio il fatturato, si può notare che la vendita media del prodotto è pari a 3.576 con una deviazione standard di 1.078.

Statistiche e grafici. Conteggi di frequenza, percentuali, percentuali cumulate, media, mediana, moda, somma, deviazione standard, varianza, intervallo, valori minimo e massimo, errore standard della media, asimmetria e curtosi (entrambe con errori standard), quartili, percentili definiti dall'utente, grafici a barre, grafici a torta e istogrammi.

Dati. Utilizzare codici numerici o stringhe corte per codificare le variabili categoriali (misure di livello nominale o ordinale).

Assunzioni. I riepiloghi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione, in particolare per le variabili con categorie ordinate o non ordinate. La maggior parte delle statistiche riassuntive, ad esempio la media e la deviazione standard, si basano sulla normale teoria e sono idonee per variabili quantitative con distribuzioni simmetriche. Le statistiche robuste, ad esempio la media, i quartili e i percentili, sono idonee per variabili quantitative rispondenti o meno all'ipotesi di normalità.

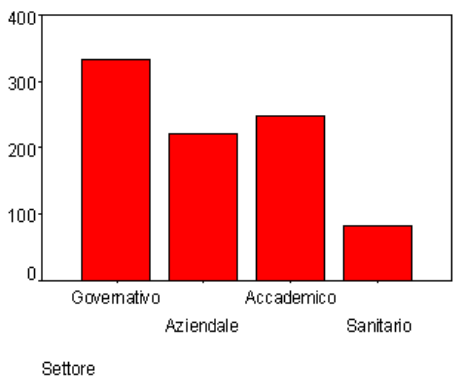
Figura 14-1
Output delle frequenze

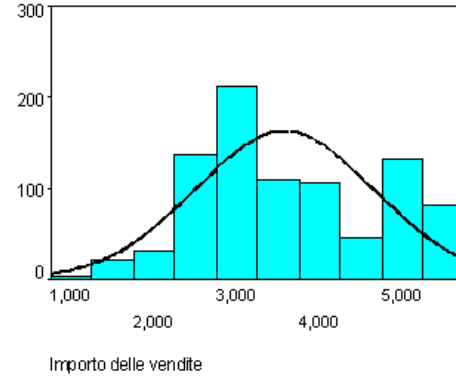
Settore

	Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Governativo	331	37.5	37.5	37.5
Aziendale	220	24.9	24.9	62.5
Accademico	248	28.1	28.1	90.6
Sanitario	83	9.4	9.4	100.0
Totale	882	100.0	100.0	

Statistiche

	Media	Mediana	Deviazione std
Importo delle vendite	$3.576,52	$3.417,50	$1.077,84





Per ottenere le tabelle di frequenza

► Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Frequenze...

Figura 14-2
Finestra di dialogo Frequenze

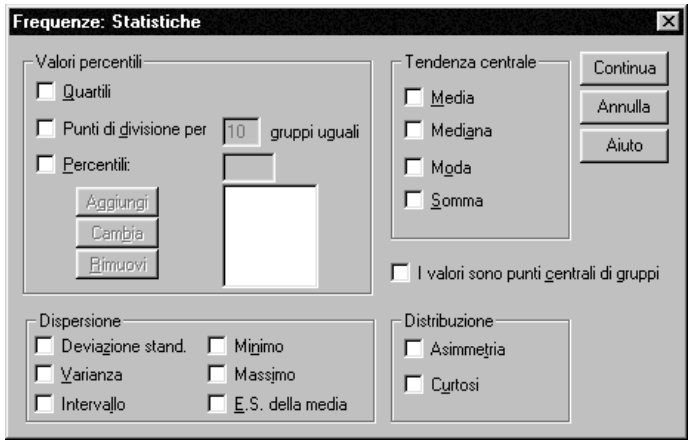
► Selezionare una o più variabili categoriali o quantitative.

Se lo si desidera, è possibile:

- Fare clic su Statistiche per ottenere statistiche descrittive per le variabili quantitative.
- Fare clic su Grafici per ottenere grafici a barre, grafici a torta e istogrammi.
- Fare clic su Formato per stabilire l'ordine in cui visualizzare i risultati.

Frequenze: Statistiche

Figura 14-3
Finestra di dialogo Frequenze: Statistiche



Valori percentili. Valori di una variabile quantitativa che suddividono i dati ordinati in due gruppi in modo da visualizzare una percentuale sopra e una sotto. I quartili (il 25°, 50° e 75° percentile) suddividono le osservazioni in quattro gruppi di dimensioni uguali. Se si desidera ottenere un numero di gruppi uguali diverso da quattro, selezionare Punti di divisione per gruppi uguali. È inoltre possibile specificare i singoli percentili, ad esempio il 95° percentile, ovvero il valore al di sotto del quale ricade il 95% delle osservazioni.

Tendenza centrale. Le statistiche che descrivono la posizione della distribuzione includono media, mediana, moda e somma di tutti i valori.

- **Media.** Una misura di tendenza centrale. La media aritmetica; la somma divisa per il numero di casi.
- **Mediana.** Una misura di tendenza centrale. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Per esempio, se ci sono 5 casi, la mediana è la terza osservazione. Quando c'è un numero pari di osservazioni, la mediana è la media delle due osservazioni centrali. La mediana non è sensibile ai valori anomali.

- **Moda.** Il valore più ricorrente. Nella serie 111234, ad esempio, la moda è 1. Se diversi valori condividono la maggiore ricorrenza, ognuno di essi è una moda. La procedura Frequenze riporta solo la più piccola delle mode.
- **Somma.** Il totale di tutti i valori validi (non mancanti) di una variabile.

Dispersione. Le statistiche che misurano l'entità della variazione o della variabilità dei dati includono deviazione standard, varianza, intervallo, valore minimo e massimo ed errore standard della media.

- **Deviazione standard.** Una misura della dispersione attorno alla media. In una distribuzione normale il 68% dei casi cade in un intervallo pari a una deviazione standard dalla media e il 95% dei casi cade in un intervallo pari a due volte la deviazione standard. Per esempio, se in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.
- **Varianza.** Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile. È il quadrato della deviazione standard
- **Intervallo.** La differenza tra il valore massimo ed il valore minimo di una variabile numerica.
- **Minimo.** Il valore più basso assunto da una variabile numerica.
- **Massimo.** Il valore più alto di una variabile numerica.
- **E. S. media.** Una misura di quanto può variare il valore della media da campione a campione per campioni estratti dalla stessa distribuzione. È la deviazione standard della distribuzione di tutte le medie possibili, se venissero estratti ripetutamente campioni della stessa dimensione. Confrontando la media osservata in un campione con un valore ipotizzato per la popolazione, si può concludere che i due valori sono significativamente diversi se il rapporto fra le differenze dei due valori e l'errore standard è maggiore di 2 in valore assoluto.

Distribuzione. L'asimmetria e la curtosi sono statistiche che descrivono la forma e la simmetria della distribuzione. Queste statistiche vengono visualizzate con i relativi errori standard.

- **Asimmetria.** Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica, e ha un valore di asimmetria pari a zero. Una distribuzione con un'asimmetria positiva significativa ha una lunga coda a destra. Una distribuzione con un'asimmetria negativa significativa ha una lunga coda a

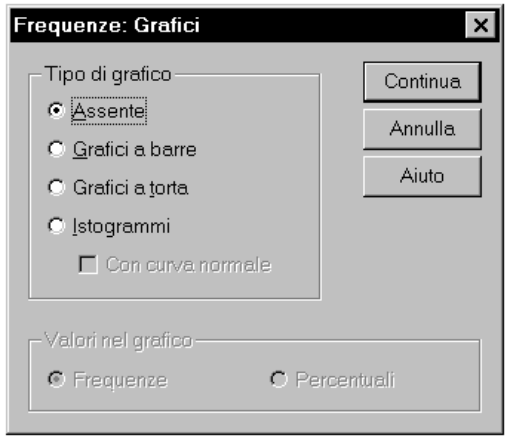
sinistra. A grandi linee, un valore di asimmetria maggiore del doppio del suo errore standard indica uno scostamento dalla simmetria.

- **Curtosi.** Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per una distribuzione normale, il valore della statistica curtosi è pari a 0. Per campioni provenienti da una distribuzione normale, i valori di curtosi oscilleranno attorno allo zero. Se una variabile ha una curtosi negativa, la sua distribuzione ha meno concentrazione nelle code rispetto ad una distribuzione normale. Se una variabile ha una curtosi positiva, una porzione di casi più elevata è concentrata nelle code rispetto alla distribuzione normale. La curtosi può essere usata, assieme all'asimmetria, per accertare che una variabile sia distribuita in modo normale.

I valori sono punti centrali di gruppi. Se i valori dei dati sono punti centrali di gruppi (ad esempio, l'età delle persone sulla trentina è codificata come 35), selezionare questa opzione per valutare la media e i percentili per i dati originali non raggruppati.

Frequenze: Grafici

Figura 14-4
Finestra di dialogo Frequenze: Grafici.



Tipo di grafico. I grafici a torta mostrano il contributo delle parti all'intero grafico. Ogni sezione di un grafico a torta corrisponde a un gruppo definito da una singola variabile di raggruppamento. Nei grafici a barre il conteggio relativo a ciascun valore o categoria viene rappresentato come una barra distinta, in modo da poter

confrontare visivamente le categorie. Anche gli istogrammi contengono barre, che però sono tracciate lungo una scala per intervalli uguali. L'altezza di ogni barra rappresenta il conteggio dei valori di una variabile quantitativa che rientra nell'intervallo. Nell'istogramma vengono indicati la forma, il centro e la variabilità della distribuzione. Una curva normale sovrapposta all'istogramma consente di valutare se i dati sono distribuiti normalmente.

Valori nel grafico. Per i grafici a barre, l'asse di scala può essere etichettato in base ai conteggi o alle percentuali di frequenza.

Frequenze: Formato

Figura 14-5
Finestra di dialogo Frequenze: Formato

Ordina per. La tabella di frequenza può essere disposta in base ai valori effettivi dei dati oppure in base al conteggio (frequenza di ricorrenza) di tali valori, in ordine crescente o decrescente. Se, tuttavia, si desidera ottenere un istogramma o i percentili, si presume che la variabile sia quantitativa e i suoi valori vengano visualizzati in ordine crescente.

Variabili multiple. Se si producono tabelle di statistiche per variabili multiple, è possibile visualizzare tutte le variabili in un'unica tabella (Confronta variabili) o visualizzare una tabella distinta di statistiche per ciascuna variabile (Output per variabili).

Sopprimi le tabelle con più di n modalità. Questa opzione consente di disattivare la visualizzazione delle tabelle che includono un numero di valori maggiore di quello specificato.

Capitolo
15

Descrittive

La procedura Descrittive consente di visualizzare statistiche riassuntive univariate per diverse variabili incluse nella stessa tabella e di calcolare i valori standardizzati (punteggi *z*). È possibile ordinare le variabili in base alle dimensioni delle rispettive medie (in ordine crescente o decrescente), in ordine alfabetico oppure nell'ordine in cui sono state selezionate (impostazione di default).

I punteggi *z* salvati vengono aggiunti ai dati nell'Editor dei dati e sono disponibili per la creazione di grafici, elenchi di dati e analisi. Quando le variabili vengono registrate in unità diverse (ad esempio, prodotto interno lordo pro capite e percentuale di alfabetizzazione), una trasformazione dei punteggi *z* consente di posizionare le variabili su una scala comune per facilitarne il confronto visivo.

Esempio. Se ciascun caso incluso nei dati contiene i totali delle vendite giornaliere relativi a ciascun agente di vendita (ad esempio, una voce per Roberto, una per Carlo, una per Bruno e così via), registrati ogni giorno per diversi mesi, la procedura Descrittive consente di calcolare la media delle vendite giornaliere per ogni agente e di ordinare i risultati dalla media di vendita maggiore alla minore.

Statistiche. Dimensioni del campione, media, valore minimo e massimo, deviazione standard, varianza, intervallo, somma, errore standard della media, curtosi e asimmetria degli errori standard.

Dati. Utilizzare variabili numeriche dopo averle valutate graficamente per registrare errori, valori anomali e anomalie distributive. La procedura Descrittive risulta molto utile quando si utilizzano file di grandi dimensioni (migliaia di casi).

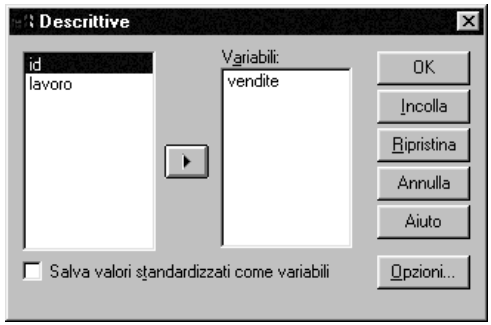
Assunzioni. La maggior parte delle statistiche disponibili (compresi i punteggi *z*) si fondano sulla teoria di normalità e possono essere utilizzate per le variabili quantitative (misurazioni a livello di intervallo o di rapporto) con distribuzioni simmetriche. Evitare variabili con categorie non ordinate o distribuzioni asimmetriche. La distribuzione dei punteggi *z* ha la stessa forma di quella dei dati originali. Pertanto, il calcolo dei punteggi *z* non rappresenta una soluzione per dati problematici.

Per ottenere statistiche descrittive

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Descrittive...

Figura 15-1
Finestra di dialogo Descrittive



▶ Selezionare una o più variabili.

Se lo si desidera, è possibile:

- Selezionare Salva valori standardizzati come variabili per salvare i punteggi *z* come nuove variabili.
- Fare clic su Opzioni per ottenere le statistiche e l'ordine di visualizzazione facoltativi.

Descrittive: Opzioni

Figura 15-2
Finestra di dialogo Descrittive: Opzioni

Media e somma. Per impostazione predefinita, viene visualizzata la media o la media aritmetica.

Dispersione. Le statistiche che misurano la dispersione o la variazione dei dati includono deviazione standard, varianza, intervallo, valore minimo e massimo ed errore standard della media.

- **Deviazione standard.** Una misura della dispersione attorno alla media. In una distribuzione normale il 68% dei casi cade in un intervallo pari a una deviazione standard dalla media e il 95% dei casi cade in un intervallo pari a due volte la deviazione standard. Per esempio, se in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.
- **Varianza.** Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile. È il quadrato della deviazione standard
- **Intervallo.** La differenza tra il valore massimo ed il valore minimo di una variabile numerica.
- **Minimo.** Il valore più basso assunto da una variabile numerica.

- **Massimo.** Il valore più alto di una variabile numerica.
- **S.E. media.** Una misura di quanto può variare il valore della media da campione a campione per campioni estratti dalla stessa distribuzione. È la deviazione standard della distribuzione di tutte le medie possibili, se venissero estratti ripetutamente campioni della stessa dimensione. Confrontando la media osservata in un campione con un valore ipotizzato per la popolazione, si può concludere che i due valori sono significativamente diversi se il rapporto fra le differenze dei due valori e l'errore standard è maggiore di 2 in valore assoluto.

Distribuzione. Curtosi e asimmetria sono statistiche che caratterizzano la forma e la simmetria della distribuzione. Queste statistiche vengono visualizzate con i relativi errori standard.

- **Curtosi.** Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per una distribuzione normale, il valore della statistica curtosi è pari a 0. Per campioni provenienti da una distribuzione normale, i valori di curtosi oscilleranno attorno allo zero. Se una variabile ha una curtosi negativa, la sua distribuzione ha meno concentrazione nelle code rispetto ad una distribuzione normale. Se una variabile ha una curtosi positiva, una porzione di casi più elevata è concentrata nelle code rispetto alla distribuzione normale. La curtosi può essere usata, assieme all'asimmetria, per accertare che una variabile sia distribuita in modo normale.
- **Asimmetria.** Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica, e ha un valore di asimmetria pari a zero. Una distribuzione con un'asimmetria positiva significativa ha una lunga coda a destra. Una distribuzione con un'asimmetria negativa significativa ha una lunga coda a sinistra. A grandi linee, un valore di asimmetria maggiore del doppio del suo errore standard indica uno scostamento dalla simmetria.

Ordine di visualizzazione. Per impostazione predefinita, le variabili vengono visualizzate nell'ordine in cui vengono selezionate. È inoltre possibile visualizzare le variabili in ordine alfabetico, per media crescente o per media decrescente.

Funzioni aggiuntive del comando DESCRIPTIVES

Il linguaggio a comandi SPSS permette anche di:

- Salvare i punteggi standardizzati (punteggi *z*) per alcune ma non per tutte le variabili (con il sottocomando `VARIABLES`).

- Specificare i nomi delle nuove variabili che contengono i punteggi standardizzati (con il sottocomando `VARIABLES`).
- Escludere dall'analisi i casi con valori mancanti per qualsiasi variabile (con il sottocomando `MISSING`).
- Ordinare le variabili visualizzate in base al valore di una statistica, non solo in base alla media (con il sottocomando `SORT`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
16

Esplora

La procedura Esplora produce statistiche riassuntive e visualizzazioni grafiche per tutti i casi o per singoli gruppi di casi. Risulta inoltre utile per numerose operazioni, ovvero screening dei dati, identificazione dei valori anomali, descrizione, verifica delle ipotesi e caratterizzazione delle differenze tra sottopopolazioni (gruppi di casi). Lo screening dei dati può evidenziare la presenza di valori insoliti, intervalli vuoti tra i dati o altri elementi specifici. L'esplorazione dei dati può consentire di determinare l'idoneità delle tecniche statistiche selezionate per l'analisi dei dati. L'esplorazione può evidenziare la necessità di eseguire una trasformazione dei dati se una particolare tecnica richiede una distribuzione normale. In alternativa è possibile utilizzare test non parametrici.

Esempio. Si consideri la distribuzione dei tempi in cui quattro gruppi di ratti imparano a uscire da un labirinto. Per ciascuno dei quattro gruppi, è possibile verificare se la distribuzione dei tempi è approssimativamente normale e se i quattro valori di varianza sono uguali. È inoltre possibile identificare i casi con i cinque tempi più lunghi e i cinque tempi più brevi. I grafici a scatole e i grafici ramo-foglia riassumono graficamente la distribuzione dei tempi di apprendimento per ciascun gruppo.

Statistiche e grafici. Media, mediana, media 5% trim, errore standard, varianza, deviazione standard, valore minimo e massimo, intervallo, distanza interquartilica, asimmetria e curtosi e i relativi errori standard, intervallo di confidenza per la media (e il livello di confidenza specificato), percentili, stimatore M di Huber, stimatore M di Andrew, stimatore M decrescente di Hampel, stimatore di Tukey a doppio peso, i cinque valori maggiori e i cinque valori minori, il test di Kolmogorov-Smirnov con il livello di significatività di Lilliefors per il test della normalità e il test di Shapiro-Wilk. Grafici a scatole, grafici ramo-foglia, istogrammi, grafici di normalità e grafici di variabilità contro intensità con test di Levene e trasformazioni.

Dati. La procedura Esplora può essere utilizzata per le variabili quantitative (livello di misurazione per intervallo o per rapporto). La variabile fattore, utilizzata per suddividere i dati in gruppi di casi, deve includere un numero ragionevole di valori distinti (categorie). Tali valori possono essere stringhe corte o numerici. La variabile etichetta, utilizzata per etichettare i valori anomali in grafici a scatole, può essere una variabile stringa corta, stringa lunga (i primi 15 caratteri) o numerica.

Assunzioni. La distribuzione dei dati non deve essere necessariamente simmetrica o normale.

Figura 16-1
Output della procedura Esplora

Descrittive

			Time			
			Schedule			
			1	2	3	4
Media		Statistica	2,760	4,850	6,900	9,010
		Errore std	,165	,422	,445	,289
Intervallo di confidenza per la media al 95%	Limite inferiore	Statistica	2,387	3,895	5,893	8,357
	Limite superiore	Statistica	3,133	5,805	7,907	9,663
5% Trimmed Mean		Statistica	2,761	4,889	6,911	8,994
Mediana		Statistica	2,850	4,900	7,050	9,000
Variance		Statistica	,272	1,783	1,982	,834
Deviazione std.		Statistica	,521	1,335	1,408	,913
Minimo		Statistica	2,0	2,3	4,5	7,8
Massimo		Statistica	3,5	6,7	9,1	10,5
Varianza		Statistica	1,5	4,4	4,6	2,7
Distanza interquartilica		Statistica	,925	2,250	2,400	1,650
Asimmetria		Statistica	-,116	-,559	-,197	,219
		Errore std	,687	,687	,687	,687
Curtosi		Statistica	-1,210	-,104	-,606	-1,350
		Errore std	1,334	1,334	1,334	1,334

Valori estremi

			Numero di caso	Programma	Valore
Tempo	Più alto	1	31	4	10.5
		2	33	4	9.9
		3	39	4	9.8
		4	32	4	9.5
		5	36	4	9.3
	Più basso	1	2	1	2.0
		2	7	1	2.1
		3	1	1	2.3
		4	11	2	2.3
		5	3	1	2.5

Frequenza Ramo-foglia

```
  7,00      2 . 0133589
  6,00      3 . 014577
  3,00      4 . 568
  5,00      5 . 05779
  4,00      6 . 1379
  3,00      7 . 268
  6,00      8 . 012237
  5,00      9 . 13589
  1,00     10 . 5
```

Larghezza ramo: 1,0
Ciasuna foglia: 1 caso/i

Per esplorare i dati

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Esplora...

Figura 16-2
Finestra di dialogo Esplora

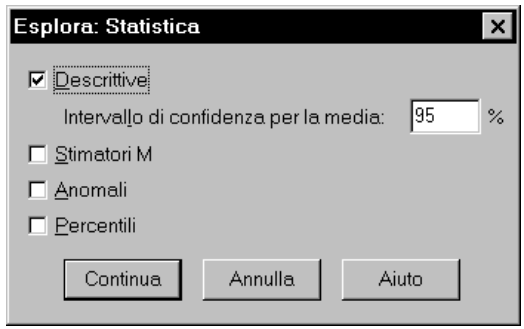
▶ Selezionare una o più variabili dipendenti.

Se lo si desidera, è possibile:

- Selezionare una o più variabili fattore i cui valori definiranno i gruppi di casi.
- Selezionare una variabile di identificazione per etichettare i casi.
- Fare clic su Statistiche per ottenere stimatori robusti, valori anomali, percentili e tabelle di frequenza.
- Fare clic su Grafici per ottenere istogrammi, grafici e test di probabilità normale e grafici di variabilità contro intensità con test di Levene.
- Fare clic su Opzioni per ottenere il trattamento dei valori mancanti.

Esplora: Statistica

Figura 16-3
Finestra di dialogo Esplora: Statistica



Descrittive. Queste misure di tendenza centrale e di dispersione vengono visualizzate per impostazione predefinita. Le misure di tendenza centrale indicano la posizione della distribuzione e includono la media, la mediana e la media 5% trim. Le misure di dispersione mostrano la dissimilarità dei valori e includono errore standard, varianza, deviazione standard, valore minimo e massimo, intervallo e distanza interquartilica. Le statistiche descrittive includono anche le misure della forma della distribuzione; l'asimmetria e la curtosi vengono visualizzate con i rispettivi errori standard. Viene visualizzato anche l'intervallo di confidenza al 95% per la media. È possibile specificare un diverso livello di confidenza.

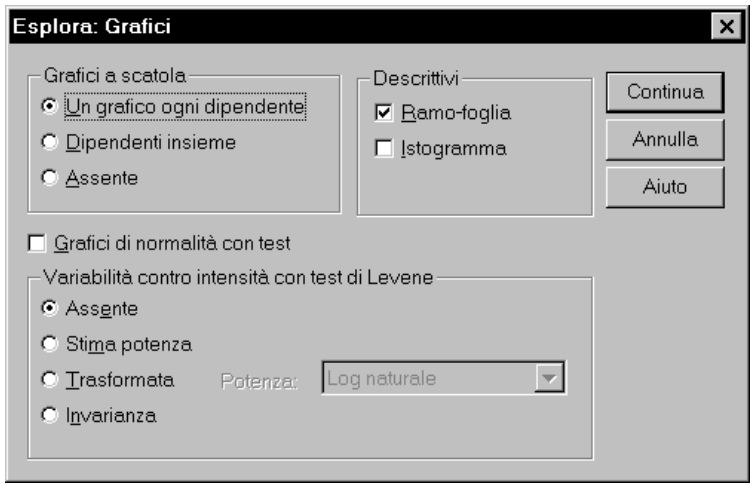
Stimatori M. Alternative valide alla media e alla mediana del campione per la valutazione della posizione. Gli stimatori calcolati differiscono per il peso applicato ai casi. Verranno visualizzati lo stimatore M di Huber, lo stimatore M di Andrews, lo stimatore M decrescente di Hampel e lo stimatore di Tukey a doppio peso.

Valori anomali. Consente di visualizzare i cinque valori maggiori e i cinque valori minori con le etichette dei casi.

Percentili. Consente di visualizzare i valori del 5°, 10°, 25°,50°, 75°, 90° e 95° percentile.

Esplora: Grafici

Figura 16-4
Finestra di dialogo Esplora: Grafici



Grafici a scatole. Queste alternative controllano la visualizzazione dei grafici a scatole quando sono presenti più variabili dipendenti. Un grafico ogni dipendente consente di generare una visualizzazione distinta per ciascuna variabile dipendente. All'interno della visualizzazione, vengono visualizzati grafici a scatole per ciascun gruppo definito da una variabile fattore. Dipendenti insieme consente di generare una visualizzazione distinta per ciascun gruppo definito da una variabile fattore. All'interno della visualizzazione compaiono grafici a scatole affiancati per ciascuna variabile dipendente. Questo tipo di grafico risulta particolarmente utile quando le singole variabili rappresentano una caratteristica misurata in tempi diversi.

Descrittive. Nel gruppo Descrittive è possibile scegliere grafici ramo-foglia e istogrammi.

Grafici di normalità con test. Consente di visualizzare grafici di probabilità normale e grafici di probabilità normale detrendizzati. Viene visualizzato il test di Kolmogorov-Smirnov con un livello di significatività di Lilliefors per il test della normalità. Se i pesi non interi sono specificati, la statistica di Shapiro-Wilk viene calcolata quando la dimensione campione pesata è compresa tra 3 e 50. Per pesi interi o non pesi, la statistica viene calcolata quando la dimensione del campione pesato è compresa tra 3 e 5000.

Variabilità vs. intensità con test di Levene. Consente di controllare la trasformazione dei dati per i grafici di variabilità contro intensità. Per tutti i grafici di variabilità contro intensità vengono visualizzati l'inclinazione della curva di regressione e i test di Levene per l'omogeneità della varianza. Se si seleziona una trasformazione, i test di Levene si baseranno sui dati trasformati. Se non viene selezionata una variabile fattore, non verranno creati grafici di variabilità contro intensità. Stima potenza traccia i logaritmi naturali delle distanze interquartiliche verso i logaritmi naturali delle mediane di tutte le celle e inoltre una stima della potenza necessaria per trasformare i dati in modo da raggiungere varianze uguali in tutte le celle. Un grafico variabilità contro intensità consente di identificare la potenza di una trasformazione per stabilizzare (rendere maggiormente uguale) le varianze nei vari gruppi. Trasformata consente di selezionare un valore di potenza alternativo, seguendo o meno le indicazioni della stima di potenza, e di produrre i grafici dei dati trasformati. La distanza interquartilica e la media dei dati trasformati verranno tracciate in un grafico. Invarianza consente di ottenere grafici relativi ai dati semplici. Equivale a una trasformazione con potenza 1.

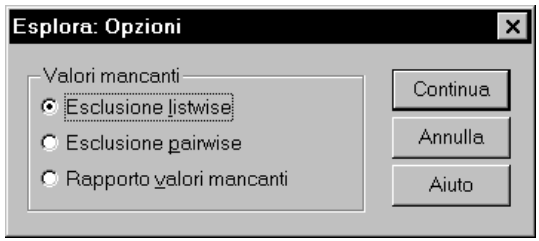
Esplora: potenza necessaria per la trasformazione dei dati

Si tratta delle trasformazioni di potenza per i grafici di variabilità contro intensità. Per trasformare i dati è necessario selezionare la potenza corrispondente. È possibile scegliere una delle seguenti opzioni:

- **Log naturale.** Trasformazione logaritmica naturale. È l'impostazione di default.
- **1/radice quadrata.** Per ciascun valore viene calcolato il reciproco della radice quadrata.
- **Reciproco.** Viene calcolato il reciproco di ciascun valore.
- **Radice quadrata.** Viene calcolata la radice quadrata di ciascun valore.
- **Quadrato.** Ciascun valore viene elevato al quadrato.
- **Cubo.** Ciascun valore viene elevato al cubo.

Esplora: Opzioni

Figura 16-5
Finestra di dialogo Esplora: Opzioni



Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile dipendente o fattore verranno esclusi da tutte le analisi. È l'impostazione di default.
- **Esclusione pairwise.** I casi che non contengono valori mancanti per le variabili di un gruppo (cella) verranno inclusi nell'analisi per tale gruppo. Il caso può includere valori mancanti per le variabili utilizzate in altri gruppi.
- **Rapporto valori mancanti.** I valori mancanti per le variabili fattore vengono trattati come categoria distinta. Tutto l'output viene prodotto per questa categoria supplementare. Le tabelle di frequenza includono categorie per i valori mancanti. I valori mancanti per una variabile fattore vengono inclusi, ma etichettati come mancanti.

Funzioni aggiuntive del comando EXAMINE

La procedura Esplora usa la sintassi del comando `EXAMINE`. Il linguaggio a comandi SPSS permette anche di:

- Richiedere l'output totale e i grafici oltre all'output e ai grafici per i gruppi definiti dalle variabili di fattore (con il sottocomando `TOTAL`).
- Specificare una scala comune per un gruppo di grafici a scatole (con il sottocomando `SCALE`).
- Specificare le interazioni delle variabili di fattore (con il sottocomando `VARIABLES`).

- Specificare percentili diversi da quelli di default (con il sottocomando `PERCENTILES`).
- Calcolare i percentili utilizzando uno dei cinque metodi (con il sottocomando `PERCENTILES`).
- Specificare una trasformazione di potenza per i grafici di variabilità vs. intensità (con il sottocomando `PLOT`).
- Specificare il numero di valori estremi da visualizzare (con il sottocomando `STATISTICS`).
- Specificare i parametri per i predittori M e i predittori robusti di posizione (con il sottocomando `MESTIMATORS`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
17

Tavole di contingenza

La procedura Tavole di contingenza consente di formare tabelle bivariate e a più dimensioni e fornisce una serie di test e misure di associazione per le tabelle bivariate. Il test o la misura da utilizzare vengono determinati in base alla struttura della tabella e al fatto che le categorie siano ordinate o meno.

Le statistiche e le misure delle tavole di contingenza vengono calcolate solo per le tabelle bivariate. Se si specifica una riga, una colonna o uno strato (variabile di controllo), verrà visualizzato un riquadro contenente le statistiche associate e le misurazioni per ciascun valore dello strato (o una combinazione di valori per due o più variabili di controllo). Ad esempio, se la variabile *sesso* è uno strato per la tabella della variabile *coniugato* (sì, no) rispetto alla variabile *tipo di vita* (ottima, soddisfacente, non soddisfacente), i risultati per la tabella bivariata per le donne vengono elaborati separatamente da quelli per gli uomini e quindi stampati come riquadri in successione.

Esempio. È possibile che i clienti rappresentati da piccole società siano più remunerativi per la vendita di servizi (per esempio addestramenti e consulenze) rispetto ai clienti rappresentati da società di grandi dimensioni? Mediante una tavola di contingenza è possibile scoprire che la maggior parte delle società di piccole dimensioni (con un numero di dipendenti inferiore a 500) fruttano alti profitti per i servizi, mentre la maggior parte delle grandi società (con oltre 2.500 dipendenti) fruttano profitti di scarsa entità.

Statistiche e misure di associazione. Chi-quadrato di Pearson, chi-quadrato del rapporto di verosimiglianza, test di associazione lineare, test esatto di Fisher, chi-quadrato corretto di Yates, *R* di Pearson , rho di Spearman, coefficiente di contingenza, phi, *V* di Cramér, lambda simmetrica e asimmetrica, tau di Goodman e Kruskal, coefficiente di incertezza, gamma, *D* di Somers, tau-*b* di Kendall, tau-*c* di Kendall, coefficiente eta, kappa di Cohen, stima del rischio relativo, rapporto odd, test di McNemar, statistiche di Cochran e Mantel-Haenszel.

Dati. Per definire le categorie di ciascuna variabile della tabella, utilizzare i valori di una variabile numerica o stringa corta (con una lunghezza massima di otto caratteri). Ad esempio, per la variabile *sesso*, è possibile codificare i dati come 1 e 2 oppure come *maschio* e *femmina*.

Assunzioni. Alcune statistiche e misure assumono categorie ordinate (dati ordinali) o valori quantitativi (dati misurati per intervallo o per rapporto), come indicato nella sezione sulle statistiche. Se le variabili della tabella prevedono categorie non ordinate (dati nominali), sono disponibili altri valori validi. Per le statistiche basate sul chi-quadrato (phi, *V* di Cramér, coefficiente di contingenza), i dati devono essere rappresentati da un campione casuale proveniente da una distribuzione multinomiale.

Nota: le variabili ordinali possono essere codici numerici che rappresentano categorie (ad esempio 1 = *basso*, 2 = *medio*, 3 = *alto*) oppure valori di stringa. Si suppone tuttavia che l'ordine alfabetico dei valori di stringa rifletta l'esatto ordine delle categorie. Ad esempio, per una variabile stringa con i valori *basso*, *medio*, *alto*, l'ordine delle categorie viene interpretato come *alto*, *basso*, *medio*, ma questo non è l'ordine corretto. In generale, per rappresentare i dati ordinali, è più sicuro utilizzare i codici numerici.

Figura 17-1
Output delle tavole di contingenza

Tavole di contingenza Redditività del servizio * Dimensione della società

Redditività del servizio	Dimensione della società			Totale
	1-500	501-2,500	> 2,500	
Min.	200	85	135	420
Max.	251	106	105	462
Totale	451	191	240	882

Test Chi-quadrato

	Valore	Df	Sig. Asint. (2 sensi)
Pearson Chi-quadrato	9,848	2	,007
Rapporto di verosimiglianza	9,852	2	,007
Lineare-lineare Associazione	7,869	1	,005
N. di casi validi	882		

Per ottenere tavole di contingenza

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Tavole di contingenza...

Figura 17-2
Finestra di dialogo Tavole di contingenza

▶ Selezionare una o più variabili di riga e una o più variabili di colonna.

Se lo si desidera, è possibile:

- Selezionare una o più variabili di controllo.
- Fare clic su Statistiche per ottenere test e misure di associazione per tabelle o sottotabelle bivariate.
- Fare clic suCelle per ottenere valori, percentuali e residui osservati e attesi.
- Fare clic suFormato per controllare l'ordine delle categorie.

Strati nelle tavole di contingenza

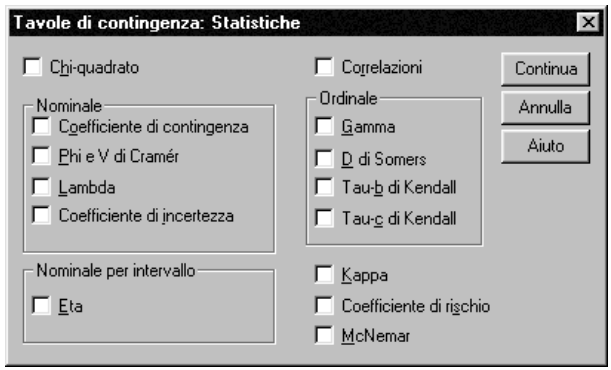
Se vengono selezionate una o più variabili di strato, verrà prodotta una tavola di contingenza distinta per ciascuna categoria di ciascuna variabile di strato (variabile di controllo). Ad esempio, se si dispone di una variabile di riga, una variabile di colonna e una variabile di strato con due categorie, si otterrà una tabella bivariata per ciascuna categoria della variabile di strato. Per creare un altro strato di variabili di controllo, fare clic su Successivo. Verranno create sottotabelle per ogni combinazione delle categorie di ciascuna variabile del primo strato con ciascuna variabile del secondo e così via. Se sono richieste statistiche e misure di associazione, verranno applicate solo alle sottotabelle bivariate.

Tavole di contingenza: Grafici a barre raggruppati

Grafici a barre raggruppati. Nei grafici a barre raggruppati è possibile riepilogare i dati relativi a gruppi di casi. È disponibile un gruppo di barre per ciascun valore della variabile specificata in Righe. La variabile che definisce le barre contenute in ogni gruppo è quella specificata in Colonne. Per ciascun valore della variabile è disponibile una serie di barre con colori e modelli diversi. Se in Colonne o Righe si specificano più variabili, verrà prodotto un grafico a barre raggruppato per ciascuna combinazione delle due variabili.

Tavole di contingenza: Statistiche

Figura 17-3
Finestra di dialogo Tavole di contingenza: Statistiche



Chi-quadrato. Per tabelle con due righe e due colonne, scegliere Chi-quadrato per calcolare il chi-quadrato di Pearson, il chi-quadrato del rapporto di verosimiglianza, il test esatto di Fisher e il chi-quadrato corretto di Yates (correzione di continuità). Per le tabelle 2 × 2, il test esatto di Fisher viene calcolato quando una tabella non creata in base a righe o colonne mancanti in una tabella di dimensioni maggiori contiene una cella con una frequenza attesa minore di 5. Per tutte le altre tabelle 2 × 2 viene calcolato il chi-quadrato corretto di Yates. Per tabelle con un numero qualsiasi di righe e colonne, selezionare Chi-quadrato per calcolare il chi-quadrato di Pearson e il chi-quadrato del rapporto di verosimiglianza. Se entrambe le variabili delle tabelle sono quantitative, l'opzione Chi-quadrato restituisce il test dell'associazione lineare.

Correlazioni. Per tabelle in cui sia le righe che le colonne contengono valori ordinati, l'opzione Correlazioni restituisce il coefficiente di correlazione di Spearman, rho (solo per dati numerici). Il coefficiente rho di Spearman è una misura di associazione tra punteggi di rango. Se entrambe le variabili delle tabelle (fattori) sono quantitative, Correlazioni restituisce il coefficiente di correlazione di Pearson, *r*, una misura dell'associazione lineare tra le variabili.

Nominale. Per i dati nominali (nessun ordine intrinseco, ad esempio cattolico, protestante, ebreo), è possibile selezionare il coefficiente Phi e V di Cramér, il Coefficiente di contingenza, Lambda (lambda simmetrico e asimmetrico e tau di Goodman e Kruskal), nonché il Coefficiente di incertezza.

- **Coefficiente di contingenza.** Una misura di associazione basata sul chi-quadrato. Questo coefficiente è sempre compreso tra 0 e 1 ma, in genere, non è possibile che assuma valore 1. Il valore massimo possibile dipende dal numero di righe e di colonne in una tabella.
- **Deviazione standard.** Una misura di dispersione intorno alla media, uguale alla radice quadrata della varianza. La deviazione standard è misurata nelle stesse unità della variabile.
- **Lambda.** Misura di associazione che riflette la riduzione nell'errore quando i valori della variabile indipendente vengono usati per stimare quelli della variabile dipendente. Un valore pari a 1 significa che la variabile indipendente stima perfettamente la variabile dipendente. Un valore pari a 0 significa che la variabile indipendente non è di nessun aiuto nella stima della variabile dipendente.
- **Coefficiente di incertezza.** Una misura della riduzione proporzionale dell'errore basata su criteri di entropia. La procedura ne calcola sia la versione simmetrica, sia quella asimmetrica. Quanto più il coefficiente di incertezza si avvicina al suo

limite superiore, pari a 1, tanto maggiori sono le informazioni sul valore della seconda variabile ottenute dalla conoscenza di un valore osservato della prima. Il suo limite inferiore è 0, quando non viene ottenuta nessuna informazione relativa alla seconda variabile da un valore osservato della prima.

Ordinale. Per tabelle in cui sia le righe che le colonne contengono valori ordinati, selezionare Gamma (gamma di ordine zero per tabelle a 2 vie e gamma condizionali per tabelle da 3 a 10 vie), Tau-b di Kendall e Tau-c di Kendall. Per desumere le categorie delle colonne delle righe, selezionare D di Somers.

- **Gamma.** Una misura simmetrica di associazione tra due variabili ordinali, che varia tra -1 ed 1 . Valori assoluti prossimi a 1 indicano una relazione più forte fra le variabili. Valori prossimi a zero indicano relazione debole o nulla. Per le tabelle a due vie, vengono visualizzati i gamma di ordine zero. Per le tabelle a tre o più vie vengono visualizzati i gamma condizionali.
- **D di Somers.** Una misura di associazione tra due variabili ordinali. Varia fra -1 e 1, dove zero indica assenza di associazione e valori prossimi a 1 in valore assoluto indicano forte relazione. È una estensione asimmetrica di gamma dalla quale differisce solo per l'inclusione del numero di coppie non a pari merito nella variabile indipendente. La D di Somers indica l'eccesso proporzionato di coppie concordanti su quelle discordanti tra coppie non a pari merito sulla variabile indipendente. La procedura calcola anche una versione simmetrica di questa statistica.
- **Tau-b di Kendall.** Una misura non parametrica di correlazione per variabili ordinali. Il segno del coefficiente indica la direzione della correlazione e il valore assoluto la sua intensità, dove valori assoluti più alti indicano relazioni più forti. I valori possibili variano da -1 a 1, ma il valore -1 o +1 può solo essere ottenuto da tabelle quadrate.
- **Tau-c di Kendall.** Misura parametrica di correlazione per variabili ordinali che ignora i valori pari merito. Il segno del coefficiente indica la direzione della correlazione e il valore assoluto la sua intensità, dove valori assoluti più alti indicano relazioni più forti. I valori possibili variano da -1 a +1, ma il valore -1 o +1 può solo essere ottenuto da tabelle quadrate.

Nominale per intervallo. Se una variabile è categoriale e l'altra quantitativa, scegliere Eta. La variabile categoriale deve essere codificata numericamente.

- **Eta.** Una misura di associazione appropriata per una variabile dipendente misurata su una scala per intervallo e una variabile indipendente con un numero limitato di categorie. Eta varia fra 0 e 1, è asimmetrica e non assume una relazione lineare tra le variabili. Il quadrato di Eta può essere interpretato come la proporzione di varianza della variabile dipendente spiegata dalle differenze tra i gruppi. Vengono calcolati due valori di Eta. Il primo assume la variabile di riga, il secondo quella di colonna, come variabile misurata su una scala per intervallo.

Kappa di Cohen. Il kappa di Cohen misura l'accordo tra due stimatori quando entrambi stanno stimando lo stesso oggetto. La differenza tra la proporzione di casi osservata nella quale gli stimatori sono in accordo e la proporzione attesa casualmente, viene divisa per la differenza massima possibile tra le proporzioni osservate ed attese, dati i totali marginali. Un valore pari a 1 indica accordo perfetto. Un valore pari a 0 indica che l'accordo può essere considerato casuale.

Rischio relativo. Nelle tabelle 2x2 il rischio relativo misura l'intensità dell'associazione tra la presenza di un fattore e la manifestazione di un evento. Se l'intervallo di confidenza della statistica include il valore 1, non si può assumere che il fattore sia associato all'evento. Il rapporto "odds ratio" può essere usato come stima del rischio relativo quando la presenza del fattore è rara.

McNemar. Test per 2 variabili dicotomiche correlate. Verifica cambiamenti di risposta usando la distribuzione chi-quadrato. Il test è molto utile in disegni sperimentali del tipo 'prima e dopo', per rilevare cambiamenti di risposta dovuti a intervento sperimentale. Per tabelle quadrate più grandi viene riportato il test di simmetria di McNemar-Bowker.

Statistiche di Cochran e Mantel-Haenszel. Le statistiche di Cochran e Mantel-Haenszel possono essere usate come test di indipendenza fra un fattore binario e una variabile di risposta binaria. Le statistiche sono corrette per modelli covariati definiti da una o più variabili di controllo.

Tavole di contingenza: Visualizzazione cella

Figura 17-4
Finestra di dialogo Tavole di contingenza: Visualizzazione cella

Per facilitare l'individuazione di modelli di dati che danno origine a un test chi-quadrato significativo, la procedura per le tavole di contingenza visualizza le frequenze attese e tre tipi di residui (devianze) che misurano la differenza tra le frequenze osservate e quelle attese. Ogni cella della tabella può contenere qualsiasi combinazione dei conteggi, delle percentuali e dei residui selezionati.

Conteggi. Il numero di casi effettivamente osservati e il numero di casi attesi se le variabili di riga e di colonna sono reciprocamente indipendenti.

Percentuali. Le percentuali possono essere aggiunte nelle righe o nelle colonne. Sono disponibili anche le percentuali del numero totale di casi rappresentati nella tabella (a strato unico).

Residui. I residui semplici non standardizzati forniscono la differenza tra valori osservati e attesi. Sono inoltre disponibili residui standardizzati e standardizzati corretti.

- **Non standardizzati.** La differenza fra un valore osservato e un valore atteso. Il valore atteso è il numero di casi che ci si aspetterebbe in una cella se non ci fossero relazioni tra le due variabili. Un residuo positivo indica che ci sono più casi nella cella di quanti ce ne sarebbero se le variabili di riga e di colonna fossero indipendenti.
- **Standardizzati.** Il residuo standardizzato, conosciuto anche come residuo di Pearson, è il residuo semplice diviso per l'errore standard della frequenza osservata. Benché i residui standardizzati siano asintoticamente normali, si

comportano in modo meno variabile (in termini di varianza) delle variabili casuali normali. La loro distribuzione ha media 0 e deviazione standard 1

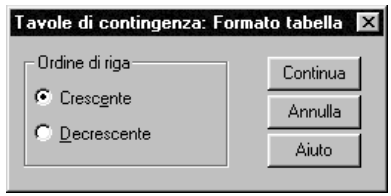
- **Standardizzati corretti.** Il residuo di una cella diviso per una stima del suo errore standard. Viene espresso in unità di deviazione standard sopra o sotto la media.

Pesi non interi. I conteggi di cella in genere sono valori interi, in quanto rappresentano il numero di casi in ogni cella. Se, tuttavia, il file di dati è attualmente ponderato in base a una variabile di ponderazione con valori frazionari (ad esempio, 1,25), i conteggi di cella possono essere espressi anche in valori frazionari. È possibile troncare o arrotondare i valori prima o dopo aver calcolato i conteggi di cella oppure utilizzare conteggi di cella frazionari per la visualizzazione delle tabelle e dei calcoli statistici.

- **Arrotonda i conteggi di cella.** I pesi dei casi sono usati come tali, ma i pesi accumulati nelle celle sono arrotondati prima di calcolare qualunque statistica.
- **Tronca i conteggi di cella.** I pesi dei casi sono usati come tali, ma i pesi accumulati nelle celle sono troncati prima di calcolare qualunque statistica.
- **Arrotonda i pesi dei casi.** I pesi dei casi sono arrotondati prima dell'uso.
- **Tronca i pesi dei casi.** I pesi dei casi sono troncati prima dell'uso.
- **Nessuna correzione.** I pesi dei casi vengono usati come sono e vengono usati conteggi di cella frazionari. Se vengono richieste statistiche esatte (disponibili solo nel modulo Exact Tests), i pesi accumulati nelle celle vengono troncati o arrotondati prima del calcolo delle statistiche esatte.

Tavole di contingenza: Formato tabella

Figura 17-5
Finestra di dialogo Tavole di contingenza: Formato tabella



È possibile disporre le righe nell'ordine crescente o decrescente dei valori della variabile di riga.

Capitolo
18

Riassumi

La procedura Riassumi consente di calcolare le statistiche di sottogruppo per le variabili all'interno delle categorie di una o più variabili di raggruppamento. Tutti i livelli della variabile di raggruppamento vengono incrociati. È possibile scegliere l'ordine in cui vengono visualizzate le statistiche. Per ciascuna variabile di tutte le categorie verranno inoltre visualizzate le statistiche riassuntive. I valori di ciascuna categoria possono essere inseriti nell'elenco o eliminati, ma negli insiemi di dati di grandi dimensioni, è possibile scegliere di elencare solo i primi *n* casi.

Esempio. Qual è l'importo medio delle vendite per area e industria del cliente? Si potrebbe scoprire che l'importo medio delle vendite è leggermente superiore nell'area occidentale rispetto alle altre aree e che ai clienti di quest'area è associato l'importo medio più alto.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del conteggio totale, percentuale della somma di gruppo, percentuale dei conteggi di gruppo, media geometrica, media armonica.

Dati. Le variabili di raggruppamento sono variabili categoriali che possono contenere valori stringa numerici o brevi. Il numero di categorie dovrebbe essere limitato. Le altre variabili dovrebbero essere classificabili.

Assunzioni. Alcune delle statistiche di sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana e l'intervallo sono statistiche robuste, idonee per le variabili quantitative che possono o meno soddisfare l'assunzione di normalità.

Figura 18-1
Output riassunto

Riepiloghi dei casi
Vendita media prodotto per area e settore industriale

	Settore industriale				
Area	Enti governativi	Imprese	Istituzioni accademiche	Sanità	Totale
Orientale	€3.594,65	€3.953,76	€3.764,91	€3.722,32	€3.735,45
Centrale	€3.370,12	€3.268,47	€3.317,81	€3.165,11	€3.305,03
Occidentale	€3.552,50	€4.649,00	€4.276,25	€4.027,00	€4.079,46
Totale	€3.503,75	€3.727,50	€3.579,76	€3.456,93	€3.576,52

Per ottenere riepiloghi dei casi

▶ Dai menu, scegliere:

Analizza
  Report
    Riepiloghi dei casi...

Figura 18-2
Finestra di dialogo Riepiloghi dei casi

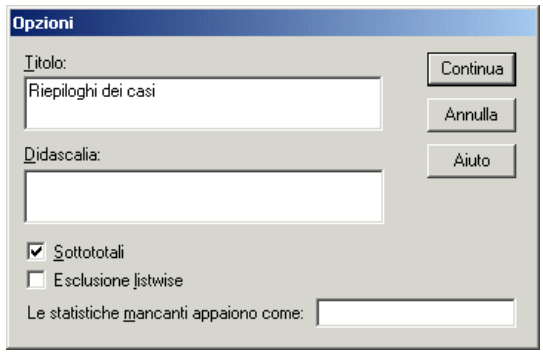
▶ Selezionare una o più variabili.

Se lo si desidera, è possibile:

- Selezionare una o più variabili di raggruppamento per suddividere i dati in sottogruppi.
- Fare clic su Opzioni per modificare il titolo dell'output, aggiungere una didascalia al di sotto dell'output o escludere casi con valori mancanti.
- Fare clic su Statistiche per visualizzare statistiche facoltative.
- Selezionare Visualizza casi per visualizzare un elenco dei casi inclusi in ciascun sottogruppo. Per impostazione predefinita, vengono elencati solo i primi 100 casi nel file. È possibile aumentare o ridurre il valore di Limita i casi ai primi*n* oppure deselezionare l'opzione per visualizzare l'elenco di tutti i casi.

Riassumi: Opzioni

Figura 18-3
Finestra di dialogo Riepiloghi dei casi: Opzioni



SPSS consente di modificare il titolo dell'output o di aggiungere una didascalia che verrà visualizzata sotto alla tabella di output. È possibile controllare gli a capo automatici dei titoli e delle didascalie digitando \n dove si desidera inserire un'interruzione di riga nel testo.

È inoltre possibile scegliere di visualizzare o eliminare i sottototali e di includere o escludere casi con valori mancanti per qualsiasi variabile utilizzata nelle analisi. È spesso consigliabile contrassegnare nell'output i casi mancanti utilizzando un punto o un asterisco. Immettere un carattere, una frase o codice che si desidera venga

visualizzato per indicare che un valore è mancante. In caso contrario, ai casi mancanti non verrà applicato alcun identificatore nell'output.

Riassumi: Statistiche

Figura 18-4
Finestra di dialogo Statistiche del comando Riepiloghi dei casi

È possibile scegliere una o più delle seguenti statistiche di sottogruppo per le variabili all'interno di ogni categoria di ciascuna variabile di raggruppamento: somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del conteggio totale, percentuale della somma di gruppo, percentuale dei conteggi di gruppo, media geometrica, media armonica. L'ordine in cui compaiono le statistiche nell'elenco Statistiche di cella corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche di riepilogo in tutte le categorie.

Primo. Visualizza il primo valore incontrato nel file dati.

Media geometrica. La radice ennesima del prodotto dei valori, dove n è il numero di casi.

Mediana dei gruppi. La mediana calcolata su valori che rappresentano gruppi (categorie distinte). Ad esempio la mediana della fascia di età.

Media armonica. Usata per stimare una dimensione media dei gruppi quando le dimensioni campionarie dei gruppi non sono uguali. La media armonica è il numero di campioni diviso per la somma dei reciproci delle dimensioni campionarie.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per una distribuzione normale, il valore della statistica curtosi è pari a 0. Per campioni provenienti da una distribuzione normale, i valori di curtosi oscilleranno attorno allo zero. Se una variabile ha una curtosi negativa, la sua distribuzione ha meno concentrazione nelle code rispetto ad una distribuzione normale. Se una variabile ha una curtosi positiva, una porzione di casi più elevata è concentrata nelle code rispetto alla distribuzione normale. La curtosi può essere usata, assieme all'asimmetria, per accertare che una variabile sia distribuita in modo normale.

Ultimo. Visualizza l'ultimo valore incontrato nel file dati.

Massimo. Il valore più alto di una variabile numerica.

Media. Una misura di tendenza centrale. La media aritmetica; la somma divisa per il numero di casi.

Mediana. Una misura di tendenza centrale. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Per esempio, se ci sono 5 casi, la mediana è la terza osservazione. Quando c'è un numero pari di osservazioni, la mediana è la media delle due osservazioni centrali. La mediana non è sensibile ai valori anomali.

Minimo. Il valore più basso assunto da una variabile numerica.

N. Numero di casi (ovvero numero di osservazioni o numero di record).

Percentuale del conteggio totale. Percentuale del numero di casi in ogni categoria.

Percentuale della somma totale. Percentuale della somma di ogni categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica, e ha un valore di asimmetria pari a zero. Una distribuzione con un'asimmetria positiva significativa ha una lunga coda a destra. Una distribuzione con un'asimmetria negativa significativa ha una lunga coda a sinistra. A grandi linee, un valore di asimmetria maggiore del doppio del suo errore standard indica uno scostamento dalla simmetria.

Errore standard della curtosi. Il rapporto fra a curtosi di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto.

Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore di asimmetria positivo ed elevato indica una lunga coda a destra, mentre un valore elevato negativo indica una lunga coda a sinistra.

Somma. Il totale di tutti i valori validi (non mancanti) di una variabile.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile. È il quadrato della deviazione standard

Capitolo
19

Medie

La procedura Medie consente di calcolare le medie dei sottogruppi e le statistiche univariate correlate per le variabili dipendenti all'interno delle categorie di una o più variabili indipendenti. È inoltre possibile ottenere analisi univariate della varianza, eta e test di linearità.

Esempio. Misurare la quantità media di grasso assorbita da tre diversi tipi di olii alimentari ed eseguire un'analisi univariata della varianza per verificare se le medie differiscono.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del conteggio totale, percentuale della somma di gruppo, percentuale dei conteggi di gruppo, media geometrica, media armonica. Le opzioni includono analisi della varianza, eta, eta quadrato, e test di linearità R e R^2.

Dati. Le variabili dipendenti sono quantitative e le variabili indipendenti sono categoriali. I valori delle variabili categoriali possono essere di tipo numerico o a stringa corta.

Assunzioni. Alcune delle statistiche di sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana è una statistica robusta, idonea per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità. L'analisi della varianza è robusta per quanto riguarda le alterazioni della normalità, ma i dati in ciascuna cella devono essere simmetrici. L'analisi della varianza assume inoltre che i gruppi provengano da popolazioni con valori di varianza uguali. Per verificare questa ipotesi, utilizzare il test di omogeneità della varianza di Levene, disponibile nella procedura ANOVA univariata.

Figura 19-1
Output delle medie

Report

Grammi di grasso assorbiti

Tipo di olio			
	Di Arachidi	Media	72.00
		N	6
		Deviazione std.	13.34
	Lardo	Media	85.00
		N	6
		Deviazione std.	7.77
	Di mais	Media	62.00
		N	6
		Deviazione std.	8.22
	Totale	Media	73.00
		N	18
		Deviazione std.	13.56

Tabella ANOVA

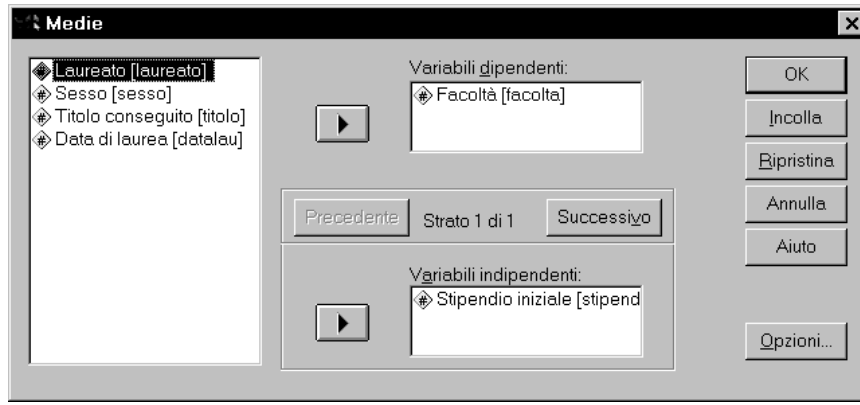
			Somma dei quadrati	df	Media dei quadrati	F	Sig.
Grammi di grasso assorbiti * Tipo di olio	Fra gruppi	(Combinata)	1596.000	2	798.000	7.824	.005
	Entro gruppi		1530.000	15	102.000		
	Totale		3126.000	17			

Per ottenere le medie dei sottogruppi

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    Medie...

Figura 19-2
Finestra di dialogo Medie



- Selezionare una o più variabili dipendenti.
- Usare uno dei seguenti metodi per selezionare le variabili categoriali indipendenti:
  - Selezionare una o più variabili indipendenti. Per ciascuna variabile indipendente vengono visualizzati risultati distinti.
  - Selezionare uno o più strati di variabili indipendenti. Ogni strato suddivide ulteriormente il campione. Se è presente una sola variabile indipendente nello Strato 1 e una sola nello Strato 2, i risultati verranno visualizzati in una tabella incrociata e non in tabelle distinte per ciascuna variabile indipendente.
- Oppure fare clic su Opzioni per ottenere statistiche facoltative, analisi della tabella di varianza, età, età quadrato, R e R^2.

Medie: Opzioni

Figura 19-3
Finestra di dialogo Medie: Opzioni

È possibile scegliere una o più delle seguenti statistiche di sottogruppo per le variabili all'interno di ogni categoria di ciascuna variabile di raggruppamento: somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero*N* totale, percentuale della somma, percentuale del *numero di casi* in, media geometrica, media armonica. È possibile modificare l'ordine in cui compaiono le statistiche per i sottogruppi. L'ordine in cui compaiono le statistiche nella lista Statistiche di cella corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche di riepilogo in tutte le categorie.

Primo. Visualizza il primo valore incontrato nel file dati.

Media geometrica. La radice ennesima del prodotto dei valori, dove n è il numero di casi.

Mediana dei gruppi. La mediana calcolata su valori che rappresentano gruppi (categorie distinte). Ad esempio la mediana della fascia di età.

Media armonica. Usata per stimare una dimensione media dei gruppi quando le dimensioni campionarie dei gruppi non sono uguali. La media armonica è il numero di campioni diviso per la somma dei reciproci delle dimensioni campionarie.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per una distribuzione normale, il valore della statistica curtosi è pari a 0. Per campioni provenienti da una distribuzione normale, i valori di curtosi oscilleranno attorno allo zero. Se una variabile ha una curtosi negativa, la sua distribuzione ha meno concentrazione nelle code rispetto ad una distribuzione normale. Se una variabile ha una curtosi positiva, una porzione di casi più elevata è concentrata nelle code rispetto alla distribuzione normale. La curtosi può essere usata, assieme all'asimmetria, per accertare che una variabile sia distribuita in modo normale.

Ultimo. Visualizza l'ultimo valore incontrato nel file dati.

Massimo. Il valore più alto di una variabile numerica.

Media. Una misura di tendenza centrale. La media aritmetica; la somma divisa per il numero di casi.

Mediana. Una misura di tendenza centrale. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Per esempio, se ci sono 5 casi, la mediana è la terza osservazione. Quando c'è un numero pari di osservazioni, la mediana è la media delle due osservazioni centrali. La mediana non è sensibile ai valori anomali.

Minimo. Il valore più basso assunto da una variabile numerica.

N. Numero di casi (ovvero numero di osservazioni o numero di record).

Percentuale del conteggio totale. Percentuale del numero di casi in ogni categoria.

Percentuale della somma totale. Percentuale della somma di ogni categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica, e ha un valore di asimmetria pari a zero. Una distribuzione con un'asimmetria positiva significativa ha una lunga coda a destra. Una distribuzione con un'asimmetria negativa significativa ha una lunga coda a sinistra. A grandi linee, un valore di asimmetria maggiore del doppio del suo errore standard indica uno scostamento dalla simmetria.

Errore standard della curtosi. Il rapporto fra a curtosi di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto.

Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore di asimmetria positivo ed elevato indica una lunga coda a destra, mentre un valore elevato negativo indica una lunga coda a sinistra.

Somma. Il totale di tutti i valori validi (non mancanti) di una variabile.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile. È il quadrato della deviazione standard

Statistiche del primo strato

Tabella Anova ed Eta. Visualizza una tabella di analisi della varianza univariata e calcola le misure di associazione eta ed eta quadrato per ogni variabile indipendente del primo strato.

Test di linearità. Calcola la somma dei quadrati degli scarti, i gradi di libertà e lo scarto quadratico medio associato con le componenti lineari e non lineari. Calcola inoltre il rappporto F, R e R quadrato. Il test viene applicato solo se la variabile indipendente è numerica.

Capitolo
20

Cubi OLAP

La procedura Cubi OLAP (Online Analytical Processing) consente di calcolare i totali, le medie e le altre statistiche univariate per le variabili riassunte continue all'interno delle categorie di una o più variabili di raggruppamento categoriali. Nella tabella viene creato uno strato distinto per ciascuna categoria di ogni variabile di raggruppamento.

Esempio. Le vendite totali e medie di diverse aree e le linee di prodotti all'interno delle aree.

Statistiche. Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero di casi totale, percentuale della somma totale entro variabili di raggruppamento, percentuale del numero di casi totale entro variabili di raggruppamento, media geometrica, media armonica.

Dati. Le variabili riassunte sono quantitative (variabili continue misurate su una scala di intervallo o di rapporto) e le variabili di raggruppamento sono categoriali. I valori delle variabili categoriali possono essere di tipo numerico o a stringa corta.

Assunzioni. Alcune delle statistiche di sottogruppo facoltative, quali la media e la deviazione standard, sono basate sulla teoria della normalità e sono idonee per le variabili quantitative con distribuzione simmetrica dei dati. La mediana e l'intervallo sono statistiche robuste, idonee per le variabili quantitative che possono o meno soddisfare l'ipotesi di normalità.

Figura 20-1
Output di Cubi OLAP

Vendite 1996 per divisione e area

Divisione : Totale
Area: Totale

Somma	$145.038.250
Media	$371.893
Mediana	$307.500
Diviazione std.	$171.311

Vendite 1996 per divisione e area

Divisione : Prodotti cliente
Area: Orientale

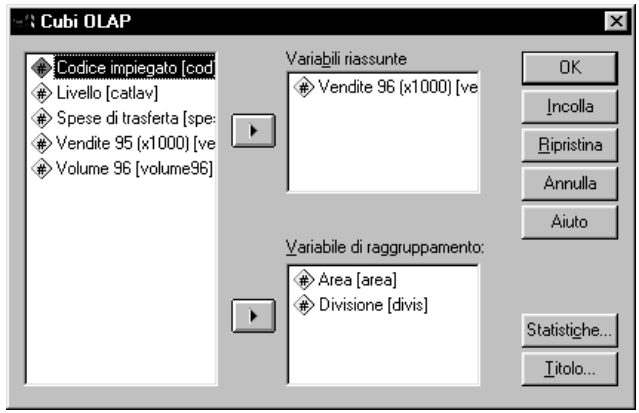
Somma	$18.548.100
Media	$289.814.06
Mediana	$273.600.00
Diviazione std.	$80.674.66

Per ottenere cubi OLAP

▶ Dai menu, scegliere:

Analizza
  Report
    Cubi OLAP...

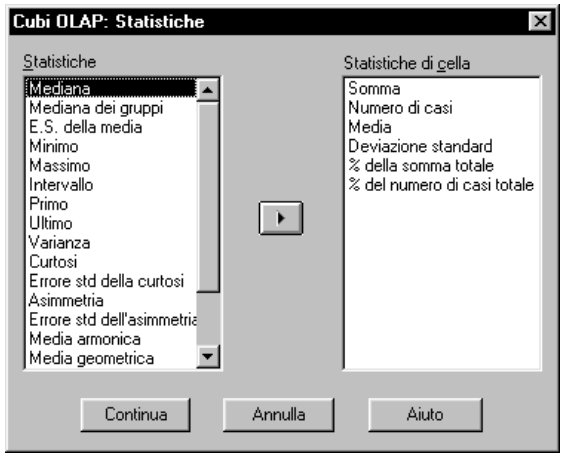
Figura 20-2
Finestra di dialogo Cubi OLAP



- Selezionare una o più variabili riassunte continue.
- Selezionare una o più variabili categoriali di raggruppamento
- Oppure:
  - Selezionare statistiche riassuntive diverse (fare clic su Statistiche). Prima di selezionare le statistiche riassuntive è necessario selezionare una o più variabili di raggruppamento.
  - Calcolare differenze tra coppie di variabili e coppie di gruppi definiti da una variabile di raggruppamento (fare clic su Differenze).
  - Creare titoli di tabella personalizzati (fare clic su Titolo).

Cubi OLAP: Statistiche

Figura 20-3
Finestra di dialogo Cubi OLAP: Statistiche



È possibile scegliere una o più delle seguenti statistiche di sottogruppo per le variabili riassunte all'interno di ogni categoria di ciascuna variabile di raggruppamento: Somma, numero di casi, media, mediana, mediana dei gruppi, errore standard della media, minimo, massimo, intervallo, valore della prima categoria della variabile di raggruppamento, valore dell'ultima categoria della variabile di raggruppamento, deviazione standard, varianza, curtosi, errore standard della curtosi, asimmetria, errore standard dell'asimmetria, percentuale della somma totale, percentuale del numero di casi totale, percentuale della somma totale entro variabili di raggruppamento, percentuale del numero di casi totale entro variabili di raggruppamento, media geometrica, media armonica.

È possibile modificare l'ordine in cui compaiono le statistiche per i sottogruppi. L'ordine in cui compaiono le statistiche nella lista Statistiche di cella corrisponde all'ordine in cui verranno visualizzate nell'output. Per ciascuna variabile vengono visualizzate anche le statistiche di riepilogo in tutte le categorie.

Primo. Visualizza il primo valore incontrato nel file dati.

Media geometrica. La radice ennesima del prodotto dei valori, dove n è il numero di casi.

Mediana dei gruppi. La mediana calcolata su valori che rappresentano gruppi (categorie distinte). Ad esempio la mediana della fascia di età.

Media armonica. Usata per stimare una dimensione media dei gruppi quando le dimensioni campionarie dei gruppi non sono uguali. La media armonica è il numero di campioni diviso per la somma dei reciproci delle dimensioni campionarie.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per una distribuzione normale, il valore della statistica curtosi è pari a 0. Per campioni provenienti da una distribuzione normale, i valori di curtosi oscilleranno attorno allo zero. Se una variabile ha una curtosi negativa, la sua distribuzione ha meno concentrazione nelle code rispetto ad una distribuzione normale. Se una variabile ha una curtosi positiva, una porzione di casi più elevata è concentrata nelle code rispetto alla distribuzione normale. La curtosi può essere usata, assieme all'asimmetria, per accertare che una variabile sia distribuita in modo normale.

Ultimo. Visualizza l'ultimo valore incontrato nel file dati.

Massimo. Il valore più alto di una variabile numerica.

Media. Una misura di tendenza centrale. La media aritmetica; la somma divisa per il numero di casi.

Mediana. Una misura di tendenza centrale. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Per esempio, se ci sono 5 casi, la mediana è la terza osservazione. Quando c'è un numero pari di osservazioni, la mediana è la media delle due osservazioni centrali. La mediana non è sensibile ai valori anomali.

Minimo. Il valore più basso assunto da una variabile numerica.

N. Numero di casi (ovvero numero di osservazioni o numero di record).

Percentuale dei conteggi di gruppo. Percentuale del numero di casi di una variabile di raggruppamento entro le categorie di altre variabili di raggruppamento. Se c'è una sola variabile di raggruppamento, questo valore è uguale alla percentuale del conteggio totale.

Percentuale della somma in. Percentuale della somma di una variabile di raggruppamento entro le categorie di altre variabili di raggruppamento. Se c'è una sola variabile di raggruppamento, questo valore è uguale alla percentuale della somma totale.

Percentuale del conteggio totale. Percentuale del numero di casi in ogni categoria.

Percentuale della somma totale. Percentuale della somma di ogni categoria.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica, e ha un valore di asimmetria pari a zero. Una distribuzione con un'asimmetria positiva significativa ha una lunga coda a destra. Una distribuzione con un'asimmetria negativa significativa ha una lunga coda a sinistra. A grandi linee, un valore di asimmetria maggiore del doppio del suo errore standard indica uno scostamento dalla simmetria.

Errore standard della curtosi. Il rapporto fra a curtosi di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto.

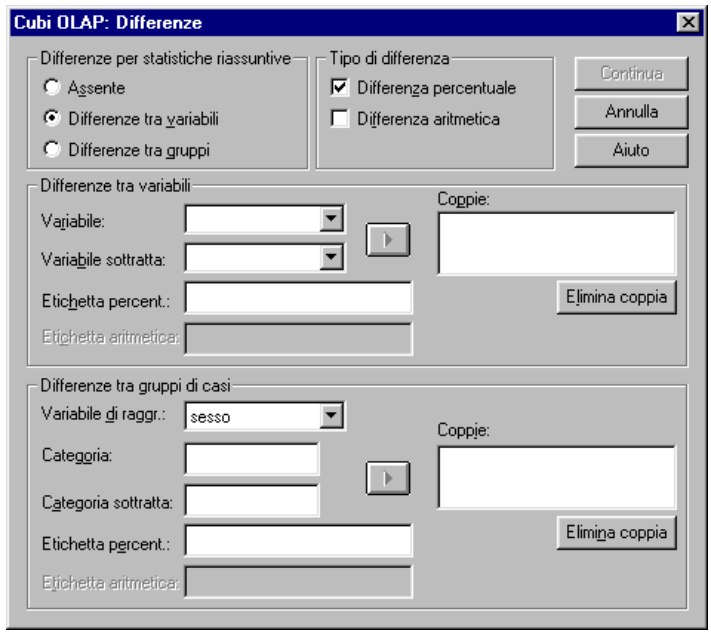
Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore di asimmetria positivo ed elevato indica una lunga coda a destra, mentre un valore elevato negativo indica una lunga coda a sinistra.

Somma. Il totale di tutti i valori validi (non mancanti) di una variabile.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile. È il quadrato della deviazione standard

Cubi OLAP: Differenze

Figura 20-4
Finestra di dialogo Cubi OLAP: Differenze



Questa finestra di dialogo consente di calcolare le differenze aritmetiche e percentuali tra variabili riassunte o tra gruppi definiti da una variabile di raggruppamento. Le differenze vengono calcolate per tutte le misure selezionate nella finestra di dialogo Cubi OLAP: Statistiche.

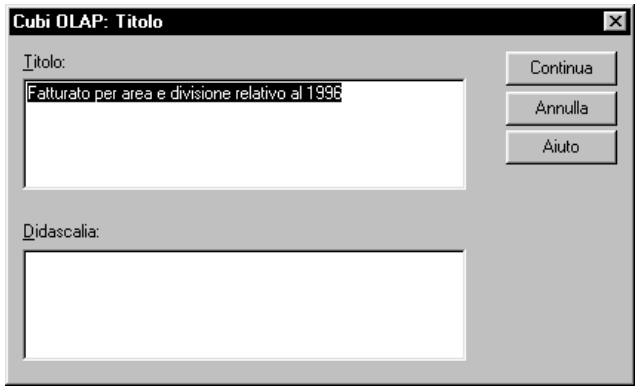
Differenze tra variabili. Consente di calcolare le differenze tra coppie di variabili. I valori delle statistiche riassuntive della seconda variabile di ogni coppia (la variabile sottratta) vengono sottratti dai valori delle statistiche riassuntive della prima variabile della coppia. Per le differenze percentuali, il valore della variabile riassunta della variabile sottratta viene utilizzato al denominatore. È necessario selezionare almeno due variabili riassunte nella finestra di dialogo principale prima di specificare le differenze tra variabili.

Differenze tra gruppi. Consente di calcolare le differenze tra coppie di gruppi definiti da una variabile di raggruppamento. I valori delle statistiche riassuntive della seconda categoria di ogni coppia (la categoria sottratta) vengono sottratti dai valori delle

statistiche riassuntive della prima categoria della coppia. Per le differenze percentuali, il valore delle statistiche riassuntive della categoria sottratta viene utilizzato al denominatore. È necessario selezionare una o più variabili di raggruppamento nella finestra di dialogo principale prima di specificare le differenze tra gruppi.

Cubi OLAP: Titolo

Figura 20-5
Finestra di dialogo Cubi OLAP: Titolo



È possibile modificare il titolo dell'output o aggiungere una didascalia che verrà visualizzata sotto la tabella dell'output. È inoltre possibile impostare gli a capo automatici dei titoli e delle didascalie digitando \n dove si desidera inserire un'interruzione di riga nel testo.

Capitolo
21

Test T

Sono disponibili tre tipi di test *t*:

Test T per campioni indipendenti (test T per due campioni). Consente di confrontare le medie di una variabile per due gruppi di casi. Vengono fornite statistiche descrittive per ciascun gruppo, il test di Levene di uguaglianza delle varianze, i valori *t* di uguaglianza e non uguaglianza della varianza e un intervallo di confidenza al 95% per la differenza tra le medie.

Test T per campioni appaiati (test T dipendente). Consente di confrontare le medie di due variabili per un singolo gruppo. Questo test viene utilizzato anche per disegni relativi a studi di confronti tra coppie o di casi di controllo. Vengono fornite statistiche descrittive per le variabili oggetto del test, la correlazione tra di esse, le statistiche descrittive per le differenze appaiate, il test *t* e un intervallo di confidenza al 95%.

Test T per un campione. Consente di confrontare la media di una variabile con un valore noto o un valore ipotizzato. Con il test *t* vengono visualizzate anche le statistiche descrittive per le variabili oggetto del test. L'output di default include un intervallo di confidenza al 95% per la differenza tra la media della variabile oggetto del test e il valore ipotizzato per il test.

Test T per campioni indipendenti

Il test T per campioni indipendenti consente di confrontare le medie relative a due gruppi di casi. Nel test, i soggetti dovrebbero essere assegnati in modo casuale a due gruppi. In questo modo, le eventuali differenze nella riposta saranno dovute alla modalità di elaborazione (o alla mancata elaborazione) e non ad altri fattori. Ciò non si verifica se si esegue il confronto tra il reddito medio di soggetti maschili e femminili. Non è infatti possibile assegnare in modo casuale una persona al sesso maschile o femminile. In questi casi, è necessario assicurarsi che le differenze relative ad altri fattori non comportino un mascheramento o l'incremento di differenze significative

nelle medie. Le differenze nel reddito medio possono essere influenzate da fattori quali il livello di educazione e non solo dal sesso al quale appartengono i soggetti.

Esempio. I pazienti con pressione sanguigna alta vengono assegnati in modo casuale a un gruppo di controllo e a un gruppo di trattamento. Ai soggetti del gruppo di controllo vengono somministrate medicine innocue e ai soggetti del gruppo trattato viene somministrato un nuovo farmaco che si ritiene possa far diminuire la pressione sanguigna. Al termine di un trattamento di due mesi, viene utilizzato il test *t* per due campioni allo scopo di confrontare i valori medi della pressione sanguigna nel gruppo di controllo e nel gruppo trattato. La pressione di ogni paziente viene misurata una volta e ciascun paziente appartiene a un solo gruppo.

Statistiche. Per ogni variabile: dimensione campione, media, deviazione standard ed errore standard della media. Per le differenze fra le medie: media, errore standard e intervallo di confidenza (è possibile specificare il livello di confidenza). Test: test di Levene di uguaglianza delle varianze e test *t* di uguaglianza delle medie per la varianza comune e la varianza separata.

Dati. I valori della variabile quantitativa desiderata si trovano in una singola colonna del file di dati. Viene utilizzata una variabile di raggruppamento che include due valori per suddividere i casi in due gruppi. La variabile di raggruppamento può essere numerica (valori quali 1 e 2, o 6,25 e 12,5) oppure una stringa breve (ad esempio *sì* e *no*). In alternativa, è possibile utilizzare una variabile quantitativa, ad esempio *età*, per suddividere i casi in due gruppi specificando un punto di divisione (il punto di divisione 21 suddivide la variabile *età* in un gruppo con meno di 21 anni e in un gruppo con più di 21 anni).

Assunzioni. Per il test *t* di uguaglianza della varianza, le osservazioni dovrebbero essere rappresentate da campioni indipendenti e casuali derivati da distribuzioni normali con la stessa varianza di popolazione. Per il test *t* di inuguaglianza della varianza, le osservazioni dovrebbero essere campioni indipendenti e casuali derivati da distribuzioni normali. Il test *t* per due campioni è sufficientemente robusto per le deviazioni dalla normalità. Durante la verifica grafica delle distribuzioni, controllare che siano simmetriche e che non siano presenti valori anomali.

Figura 21-1
Output del test T per campioni indipendenti

Statistiche di gruppo

			N	Media	Deviazione std.	Errore std. Media
Pressione sanguigna	Trattamento	Placebo	96	,77	,423	,043
		Medicina	104	,76	,429	,042

Test per campioni indipendenti

		Test di Levene di uguaglianza delle varianze		Test t di uguaglianza delle medie						
									Intervallo di confidenza per la differenza al 95%	
		F	Sig.	t	df	Sig. (2-code)	Differenza fra medie	Differenza errore standard	Inferiore	Superiore
Pressione sanguigna	Assumi varianze uguali	,094	,759	,048	198	,962	,005	,100	-,193	,202
	Non assumere varianze uguali			,048	197,259	,962	,005	,100	-,193	,202
Pressione sanguigna	Assumi varianze uguali	,181	,671	-,216	198	,829	-,015	,071	-,154	,124
	Non assumere varianze uguali			-,216	196,594	,830	-,015	,071	-,154	,124

Per ottenere un test T per campioni indipendenti

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    Test T: campioni indipendenti...

Figura 21-2
Finestra di dialogo Test T per campioni indipendenti

▶ Selezionare una o più variabili quantitative oggetto del test. Per ciascuna variabile viene calcolato un test *t* distinto.

- Selezionare una variabile di raggruppamento singola e fare clic su Definisci gruppi per specificare due codici per i gruppi che si desidera confrontare.
- Se necessario, fare clic su Opzioni per verificare in quale modo vengono considerati i dati mancanti e il livello dell'intervallo di confidenza.

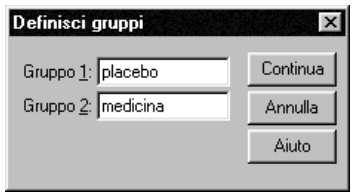
Test T per campioni indipendenti: Definisci gruppi

Figura 21-3
Finestra di dialogo Definisci gruppi per le variabili numeriche

Per le variabili di raggruppamento numeriche, definire i due gruppi per il test *t* specificando due valori o un punto di divisione:

- **Usa i valori specificati.** Inserire un valore per Gruppo 1 e un altro valore per Gruppo 2. I casi con qualsiasi altro valore verranno esclusi dall'analisi. Non è necessario specificare numeri interi (ad esempio, 6,25 e 12,5 sono validi).
- **Punto di divisione.** Immettere un numero che suddivide i valori della variabile di raggruppamento in due insiemi Assegna i casi con valori minori o uguali a quello specificato a un gruppo e i casi con valori maggiori all'altro.

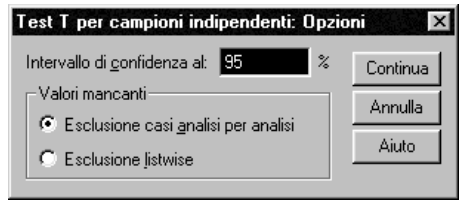
Figura 21-4
Finestra di dialogo Definisci gruppi per le variabili stringa



Per le variabili stringa di raggruppamento brevi, inserire una stringa per Gruppo 1 e un'altra stringa per Gruppo 2, ad esempio *sì* e *no*. I casi che includono altre stringhe verranno esclusi dall'analisi.

Test T per campioni indipendenti: Opzioni

Figura 21-5
Finestra di dialogo Test T per campioni indipendenti: Opzioni



Intervallo di confidenza. Per default, viene visualizzato un intervallo di confidenza del 95% per la differenza fra le medie. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Esclusione casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per le variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Esclusione listwise.** Per ciascun test *t* vengono utilizzati solo i casi con dati validi per tutte le variabili prese in considerazione nei test *t*. La dimensione del campione è costante nei vari test.

Test T per campioni appaiati

La procedura Test T per campioni appaiati consente di confrontare le medie di due variabili per un singolo gruppo. La procedura calcola le differenze tra i valori delle due variabili per ciascun caso e viene verificato se la media è diversa da 0.

Esempio. In uno studio su pazienti con valori elevati della pressione sanguigna, a tutti i pazienti è stata misurata la pressione all'inizio dello studio, è stato somministrato un trattamento e quindi la misurazione è stata ripetuta. Per ciascun soggetto sono quindi disponibili due misurazioni, in genere denominate *precedente* e *successiva*. Questo test viene utilizzato anche per disegni relativi a studi di confronti tra coppie o di casi di controllo, in cui ciascun record del file di dati contiene la risposta per il paziente e quella del soggetto di controllo corrispondente. In uno studio sulla pressione sanguigna, è necessario che l'età dei pazienti trattati corrisponda a quella

dei controlli (a un paziente di 75 anni deve corrispondere un membro del gruppo di controllo di 75 anni).

Statistiche. Per ogni variabile: media, dimensione campione, deviazione standard ed errore standard della media. Per ciascuna coppia di variabili: correlazione, differenza media tra le medie, test *t* e intervallo di confidenza per la differenza tra medie (è possibile specificare il livello di confidenza). Deviazione standard ed errore standard della differenza fra medie.

Dati. Per ciascun test appaiato, specificare due variabili quantitative (livello di misura in base a intervallo o a rapporto). In uno studio di confronti tra coppie o di casi di controllo, la risposta per ciascun soggetto del test e per il soggetto di controllo corrispondente deve trovarsi nello stesso caso all'interno del file di dati.

Assunzioni. Le osservazioni per ciascuna coppia devono essere effettuate nelle medesime condizioni. Le differenze tra medie devono essere distribuite normalmente. Le varianze di ciascuna variabile possono essere uguali o non uguali.

Figura 21-6
Output del test T per campioni appaiati

Statistiche campioni appaiati

		Media	N	Deviazione Std.	Errore std. Media
Coppia 1	Prima del Trattamento	116.40	10	13.62	4.31
	Dopo il Trattamento	142.50	10	17.04	5.39

Test per campioni appaiati

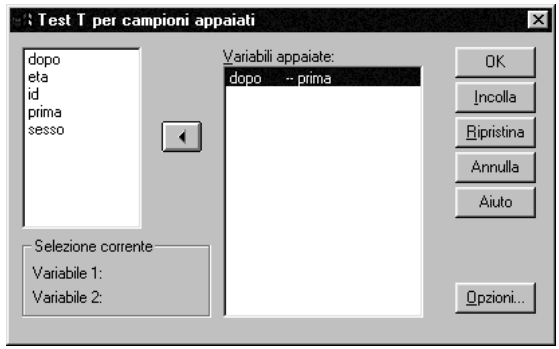
		Differenze a coppie					t	df	Sig. (2-code)
		Media	Deviazione std.	Media errore std.	Intervallo di confidenza per la differenza al 95% Inferiore	Superiore			
Coppia 1	Dopo il trattamento - Prima del trattamento	-26.10	19.59	6.19	-40.11	-12.09	-4.214	9	.002

Per ottenere un test T per campioni appaiati

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    Test T per campioni appaiati...

Figura 21-7
Finestra di dialogo Test T per campioni appaiati



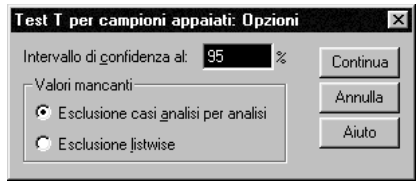
▶ Selezionare una coppia di variabili, nel modo seguente:

- Fare clic su ciascuna variabile. La prima variabile comparirà nel gruppo Selezione corrente come *Variabile 1*, mentre la seconda comparirà come *Variabile 2*.
- Fare clic sul pulsante freccia per spostare la coppia nell'elenco Variabili appaiate. È possibile selezionare più coppie di variabili. Per rimuovere una coppia di variabili dall'analisi, selezionare la coppia nella lista Variabili appaiate e fare clic sul pulsante freccia.

▶ Se necessario, fare clic su Opzioni per verificare in quale modo vengono considerati i dati mancanti e il livello dell'intervallo di confidenza.

Test T per campioni appaiati: Opzioni

Figura 21-8
Finestra di dialogo Test T per campioni appaiati: Opzioni



Intervallo di confidenza. Per default, viene visualizzato un intervallo di confidenza del 95% per la differenza fra le medie. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Esclusione casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per la coppia di variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Esclusione listwise.** Per ciascun test *t* vengono utilizzati solo i casi che includono dati validi per tutte le coppie di variabili verificate. La dimensione del campione è costante nei vari test.

Test T per un campione

La procedura Test T per un campione consente di verificare se la media di una singola variabile è diversa da una costante specificata.

Esempio. Un ricercatore può verificare se il quoziente d'intelligenza (QI) medio di un gruppo di studenti è diverso da 100. Oppure, un produttore di cereali può prelevare un campione di scatole dalla linea di produzione e verificare se il peso medio dei campioni è diverso da 0,7 kg con un livello di confidenza al 95%.

Statistiche. Per ciascuna variabile oggetto del test: media, deviazione standard ed errore standard della media. La differenza media fra ciascun valore e il valore oggetto del test ipotizzato, un test *t* che verifica che la differenza sia uguale a 0 e un intervallo di confidenza per la differenza (è possibile specificare il livello di confidenza).

Dati. Per confrontare i valori di una variabile quantitativa con un valore oggetto del test ipotizzato, scegliere una variabile quantitativa e immettere un valore oggetto del test ipotizzato.

Assunzioni. In questo test si presume che i dati siano distribuiti in modo normale. Il test è tuttavia sufficientemente robusto per le deviazioni dalla normalità.

Figura 21-9
Output del test T per un campione

Statistiche per un campione

	IQ
N	15
Media	109,33
Deviazione std.	12,03
Errore std. Media	3,11

Righe e colonne sono state trasposte.

Test per un campione

	Valore test = 100					
					Intervallo di confidenza del 95% per la differenza	
	t	df	Significatività (2 code)	Differenza Media	Inferiore	Superiore
QI	3,005	14	,009	9,33	2,67	15,99

Per ottenere un test T per un campione

► Dai menu, scegliere:

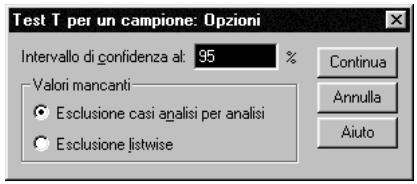
Analizza
Confronta medie
Test T campione unico...

Figura 21-10
Finestra di dialogo Test T per un campione

- Selezionare una o più variabili da confrontare con lo stesso valore ipotizzato.
- Immettere un valore oggetto del test numerico rispetto al quale viene confrontata ciascuna media del campione.
- Se necessario, fare clic su Opzioni per verificare in quale modo vengono considerati i dati mancanti e il livello dell'intervallo di confidenza.

Test T per un campione: Opzioni

Figura 21-11
Finestra di dialogo Test T per un campione: Opzioni



Intervallo di confidenza. Per default, viene visualizzato un intervallo di confidenza del 95% per la differenza fra la media e il valore oggetto del test ipotizzato. Immettere un valore compreso fra 1 e 99 per richiedere un livello di confidenza differente.

Valori mancanti. Se durante un test su più variabili si riscontra in alcune di esse la presenza di dati mancanti, è possibile indicare alla procedura i casi da includere (o da escludere):

- **Esclusione casi analisi per analisi.** Per ciascun test *t* vengono utilizzati tutti i casi con dati validi per le variabili verificate. Le dimensioni del campione possono variare in base al test.
- **Esclusione listwise.** Per ciascun tes*t* vengono utilizzati solo casi con dati validi per tutte le variabili prese in considerazione nei *t* test richiesti. La dimensione del campione è costante nei vari test.

Opzioni aggiuntive del comando T-TEST

Il linguaggio a comandi SPSS permette anche di:

- Effettuare test di T per un campione e per campioni indipendenti tramite un unico comando.
- Confrontare ciascuna variabile con le variabili dell'elenco in test accoppiati (con il sottocomando `PAIRS`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
22

ANOVA univariata

La procedura ANOVA univariata produce un'analisi della varianza univariata per una variabile dipendente quantitativa in base a una singola variabile fattore (indipendente). L'analisi della varianza consente di verificare l'ipotesi di uguaglianza di più medie. Questa tecnica è un'estensione del test *t* per due campioni.

Oltre a determinare le differenze tra le medie, è possibile individuare la media che differisce dalle altre. Esistono due tipi di test per il confronto tra le medie: contrasti a priori e test post hoc. I contrasti sono test impostati *prima* di eseguire l'esperimento, mentre i test post hoc vengono effettuati *dopo* l'esecuzione dell'esperimento. È inoltre possibile verificare i trend presenti tra le categorie.

Esempio. Le ciambelle assorbono quantità variabili di grassi a seconda della modalità di cottura. È stato condotto un esperimento che prevede l'utilizzo di tre tipi di grassi: olio di semi di arachide, olio di mais e strutto. L'olio di semi di arachide e l'olio di mais sono grassi insaturi, mentre lo strutto è un grasso saturo. Oltre a determinare se la quantità di grassi assorbita dipende dal tipo di grasso utilizzato, è possibile impostare un contrasto a priori per determinare se la quantità di grassi assorbita differisce per i grassi saturi e insaturi.

Statistiche. Per ciascun gruppo: numero di casi, media, deviazione standard, errore standard della media, valore minimo e massimo e intervallo di confidenza al 95% per la media. Test di omogeneità della varianza di Levene, tabella di analisi della varianza e test robusti dell'uguaglianza delle medie per ciascuna variabile dipendente, contrasti a priori definiti dall'utente e test di intervallo post hoc e confronti multipli: Bonferroni, Sidak, differenze significative di Tukey, GT2 di Hochberg, Gabriel, Dunnett, procedura di Ryan-Einot-Gabriel-Welsch basata su test *F*, (R-E-G-W *F*), test a intervalli di Ryan-Einot-Gabriel-Welsch (R-E-G-W *Q*), T2 di Tamhane, T3 di Dunnett, Games-Howell, *C* di Dunnett, test a intervalli multipli di Duncan, Student-Newman-Keuls (S-N-K), *b*di Tukey, Waller-Duncan, Scheffé differenza meno significativa.

Dati. I valori delle variabili fattore devono essere interi e la variabile dipendente deve essere quantitativa (livello di misura per intervallo).

Assunzioni. Ciascun gruppo è un campione casuale indipendente prelevato da una popolazione normale. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. I gruppi devono provenire da popolazioni con varianze uguali. Per verificare questa ipotesi, utilizzare il test dell'omogeneità della varianza di Levene.

Figura 22-1
Output dell'ANOVA univariata

ANOVA

		Somma dei quadrati	df	Media dei quadrati	F	Sig.
Grammi di grasso assorbiti	Entro gruppi	1596.000	2	798.000	7.824	.005
	Entro gruppi	1530.000	15	102.000		
	Totale	3126.000	17			

Descrittive

			N	Media	Deviazione std.	Errore std.	Intervallo di confidenza 95% per la media		Minimo	Massimo
							Limite inferiore	Limite superiore		
Grammi di grasso assorbiti	Tipo di olio	Di Arachidi	6	72.00	13.34	5.45	58.00	86.00	56	95
		Lardo	6	85.00	7.77	3.17	76.84	93.16	77	97
		Di mais	6	62.00	8.22	3.36	53.37	70.63	49	70
		Totale	18	73.00	13.56	3.20	66.26	79.74	49	97

Coefficienti di contrasto

	Tipo di olio		
Contrasto	Di Arachidi	Lardo	Di mais
1	-.5	1	-.5

Test di contrasto

	Grammi di grasso assorbiti	
	Assumi varianze uguali	Non presume varianze uguali
	Contrasto	Contrasto
	1	1
Valore di contrasto	18.00	18.00
Errore std.	5.05	4.51
t	3.565	3.995
df	15	12.542
Sig. (2-code)	.003	.002

Test di omogeneità delle varianze

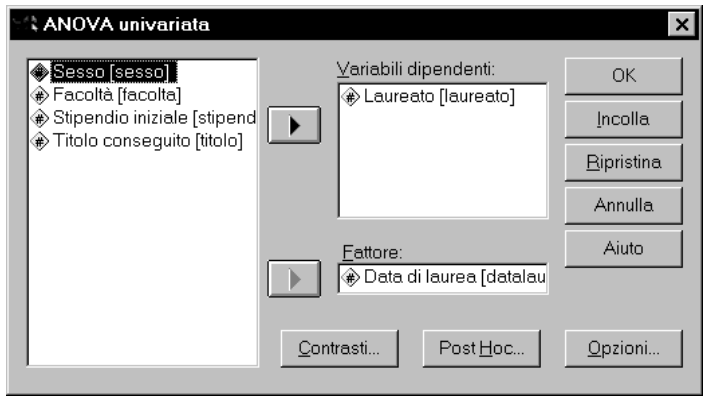
	Statistica di Levene	df1	df2	Sig.
Grammi di grasso assorbiti	.534	2	15	.597

Per ottenere un'analisi della varianza univariata

▶ Dai menu, scegliere:

Analizza
  Confronta medie
    ANOVA univariata...

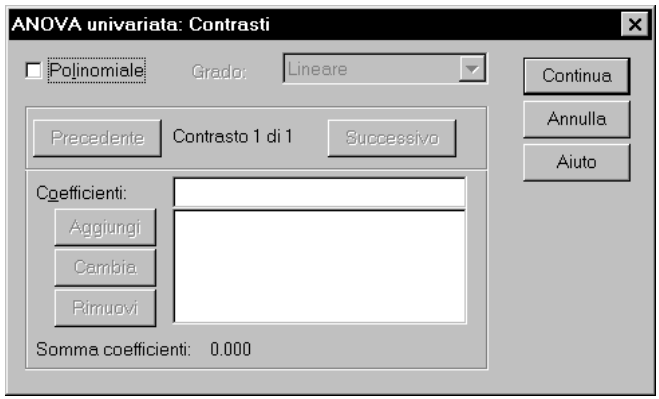
Figura 22-2
Finestra di dialogo ANOVA univariata



- Selezionare una o più variabili dipendenti.
- Selezionare una singola variabile fattore indipendente.

ANOVA univariata: Contrasti

Figura 22-3
Finestra di dialogo ANOVA univariata: Contrasti



È possibile suddividere le somme dei quadrati fra gruppi in componenti di trend oppure specificare contrasti a priori.

Polinomiale. Consente di suddividere le somme dei quadrati tra gruppi in componenti di trend. È possibile verificare un trend della variabile dipendente in tutti i livelli ordinati della variabile fattore. Ad esempio, è possibile verificare un trend lineare (crescente o decrescente) nei salari in tutti i livelli ordinati del grado di salario più elevato.

- **Grado.** È possibile scegliere un termine polinomiale di ordine 1, 2, 3, 4 o 5.

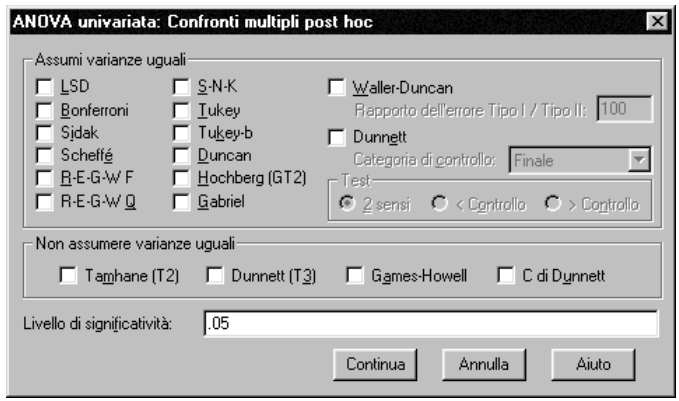
Coefficienti. Contrasti a priori definiti dall'utente da verificare con la statistica *t*. Specificare un coefficiente per ciascun gruppo (categoria) della variabile fattore e quindi fare clic su Aggiungi dopo aver inserito ciascuna voce. I nuovi valori verranno aggiunti alla fine dell'elenco dei coefficienti. Per specificare altri insiemi di contrasti, fare clic su Successivo. Utilizzare Successivo e Precedente per spostarsi tra gli insiemi di contrasti.

L'ordine dei coefficienti è importante in quanto corrisponde all'ordine crescente dei valori delle categorie della variabile fattore. Il primo coefficiente dell'elenco corrisponde al valore di gruppo minimo della variabile fattore e l'ultimo coefficiente

corrisponde al valore massimo. Ad esempio, se sono presenti sei categorie della variabile fattore, i coefficienti –1, 0, 0, 0, 0,5 e 0,5 contrastano il primo gruppo con il quinto e il sesto gruppo. Per la maggior parte delle applicazioni la somma dei coefficienti deve essere 0. È possibile utilizzare anche insiemi la cui somma è diversa da 0, ma verrà visualizzato un messaggio di avvertimento.

ANOVA univariata: Test Post Hoc

Figura 22-4
Finestra di dialogo ANOVA univariata: Confronti multipli Post Hoc



Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervalli e i confronti a coppie multipli consentono di determinare quale media differisce dalle altre. I test a intervalli multipli consentono di identificare sottoinsiemi omogenei di medie che non differiscono le une dalle altre. Grazie ai confronti a coppie multipli è possibile verificare la differenza tra ciascuna coppia di medie e ottenere una matrice in cui gli asterischi indicano le medie di gruppo con differenze significative e un livello alfa 0,05.

Assumi varianze uguali

Il test delle differenze significative di Tukey, il GT2 di Hochberg, il test di Gabriel e il test di Scheffé sono test a confronti e intervalli multipli. Sono disponibili altri test di intervalli, ovvero *b* di Tukey, S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W *F* (test *F* di Ryan-Einot-Gabriel-Welsch), R-E-G-W *Q* (test a intervalli di Ryan-Einot-Gabriel-Welsch) e Waller-Duncan. I test a confronti multipli disponibili

sono i seguenti: Bonferroni, test delle differenze significative di Tukey, Sidak, Gabriel, Hochberg, Dunnett, Scheffé e LSD (differenza meno significativa).

- **LSD.** Usa i test t per eseguire tutti i confronti a coppie tra medie di gruppo. Non viene apportata alcuna correzione al tasso di errore per i confronti multipli.
- **Bonferroni.** Usa i test t per eseguire confronti a coppie tra medie di gruppo, ma controlla il tasso di errore globale impostando il tasso di errore di ogni test al tasso di errore sperimentale diviso per il numero totale dei test. Così il livello di significatività osservato è corretto tenendo conto che si stanno effettuando confronti multipli.
- **Sidak.** Test per confronti a coppie multipli basato sul test t. Effettua la correzione del livello di significatività per confronti multipli e fornisce una banda più stretta rispetto al test di Bonferroni.
- **Scheffé.** Effettua confronti congiunti a coppie simultanei per tutte le possibili coppie di medie. Usa la distribuzione campionaria F. Può essere usato per esaminare tutte le possibili combinazioni lineari di medie di gruppo, non solo i confronti a coppie.
- **R-E-G-W F.** La procedura di Ryan-Einot-Gabriel-Welsch, basata su un test F.
- **R-E-G-W Q.** La procedura di Ryan-Einot-Gabriel-Welsch, basata su un intervallo studentizzato.
- **S-N-K.** Effettua tutti i confronti a coppie fra medie usando la distribuzione studentizzata. Per dimensioni campionarie uguali, confronta anche coppie di medie entro sottoinsiemi omogenei, utilizzando una procedure per passi. Le medie vengono ordinate dalla più grande alla più piccola, e le differenze estreme vengono verificate per prime.
- **Tukey.** Usa la statistica di intervallo studentizzato per effettuare tutti i confronti a coppie tra gruppi. Imposta il tasso di errore sperimentale al valore del tasso di errore per l'insieme di tutti i confronti per coppie.
- **b di Tukey.** Utilizza la distribuzione di intervallo studentizzato per effettuare confronti a coppie tra gruppi. Il valore critico è la media fra il corrispondente valore per il test HSD di Tukey e quello del test di Student-Newman-Keuls.
- **Duncan.** Effettua confronti a coppie usando lo stesso ordine di confronti per passi usato nel test di Student Newman Keuls, impostando un livello di soglia per il tasso di errore valido per l'insieme dei test, invece di un tasso di errore diverso per ciascun test. Usa la statistica dell'intervallo studentizzato.

- **Hochberg (GT2).** Test per confronti multipli o per intervallo basato sul modulo studentizzato. Simile al test HSD di Tukey.
- **Gabriel.** Test per confronti a coppie basato sul modulo studentizzato, generalmente più indicato del test GT2 quando le celle hanno dimensioni diverse. Può diventare poco conservativo se la variabilità delle dimensioni delle celle è molto alta.
- **Waller-Duncan.** Test per confronti multipli basato su un test t. Utilizza un approccio bayesiano.
- **Dunnett.** Test per confronti a coppie multipli basato sul test t. Confronta le medie ai diversi livelli del fattore con una media di controllo. L'ultima categoria è la categoria di controllo di default. In alternativa, è possibile scegliere la prima categoria. I test a due sensi consentono di verificare che la media in qualsiasi livello del fattore (ad eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo. I test di <controllo consentono di verificare se la media di qualsiasi livello del fattore sia minore di quella della categoria di controllo. I test di controllo > consentono di verificare se la media di qualsiasi livello del fattore sia maggiore di quella della categoria di controllo.

Non assumere varianze uguali

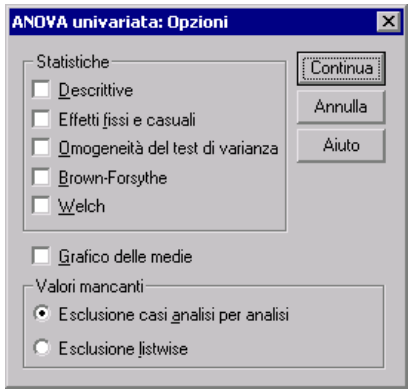
Sono disponibili test a confronti multipli che non ipotizzano varianze uguali, ovvero T2 di Tamhane, T3 di Dunnett, Games-Howell e *C* di Dunnett.

- **Tamhane (T2).** Test per confronti a coppie basato sul test t, appropriato quando le varianze non sono uguali.
- **Dunnett (T3).** Test per confronti a coppie basato sul modulo studentizzato, appropriato quando le varianze non sono uguali.
- **Games-Howell.** Test per confronti a coppie non molto conservativo, appropriato quando le varianze non sono uguali.
- **C di Dunnett.** Test per confronti a coppie basato sull'intervallo studentizzato, appropriato quando le varianze non sono uguali.

Nota: per interpretare più rapidamente l'output dei test post hoc è consigliabile deselezionare Nascondi righe e colonne vuote nella finestra di dialogo Proprietà tabella (in una tabella pivot attivata, scegliere Proprietà tabella dal menu Formato).

ANOVA univariata: Opzioni

Figura 22-5
Finestra di dialogo ANOVA univariata: Opzioni



Statistiche. Consente di scegliere una o più delle seguenti opzioni:

- **Descrittive.** Consente di calcolare il numero di casi, la media, la deviazione standard, l'errore standard della media, il valore minimo e massimo e gli intervalli di confidenza al 95% per ciascuna variabile dipendente di ciascun gruppo.
- **Effetti fissi e casuali.** Consente di visualizzare la deviazione standard, l'errore standard e l'intervallo di confidenza del 95% per il modello degli effetti fissi e l'errore standard, l'intervallo di confidenza del 95% e la stima della varianza tra componenti per il modello degli effetti casuali.
- **Omogeneità del test di varianza.** Consente di calcolare il test di Levene per verificare l'uguaglianza tra gruppi di variabili. Questo test non si basa sull'ipotesi di normalità.
- **Brown-Forsythe.** Consente di calcolare la statistica di Brown-Forsythe per verificare l'uguaglianza delle medie dei gruppi. Questa statistica è preferibile alla statistica *F* nel caso in cui non sia valida l'ipotesi di uguaglianza della varianza.
- **Welch.** Consente di calcolare la statistica di Welch per verificare l'uguaglianza delle medie dei gruppi. Questa statistica è preferibile alla statistica *F* nel caso in cui non sia valida l'ipotesi di uguaglianza della varianza.

Grafico delle medie. Consente di visualizzare un grafico che rappresenta le medie dei sottogruppi (le medie di ciascun gruppo definite dai valori della variabile fattore).

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi analisi per analisi.** I casi con valori mancanti per la variabile dipendente o fattore per una particolare analisi non verranno utilizzati in tale analisi. Non verranno utilizzati nemmeno i casi che non rientrano nell'intervallo specificato per la variabile fattore.
- **Esclusione listwise.** I casi con valori mancanti per la variabile fattore o qualsiasi variabile dipendente inclusa nella lista delle variabili dipendenti nella finestra di dialogo principale verranno esclusi da tutte le analisi. Se non sono state specificate più variabili dipendenti, l'opzione non produrrà alcun effetto.

Opzioni aggiuntive del comando ONEWAY

Il linguaggio a comandi SPSS permette anche di:

- Ottenere statistiche con effetti fissi e casuali. Deviazione standard, errore standard della media e intervalli di confidenza al 95% per il modello con effetti fissi. Errore standard, intervalli di confidenza al 95% e stima della varianza dei componenti per il modello con effetti casuali (con il sottocomando `STATISTICS=EFFECTS`).
- Specificare i livelli alfa per la differenza meno significativa nei test per confronti a coppie multipli di Bonferroni, Duncan e Scheffé (con il sottocomando `RANGES`).
- Scrivere una matrice di medie, deviazioni standard e frequenze, oppure leggere una matrice di medie, frequenze, varianze raggruppate e gradi di libertà per le varianze raggruppate. Queste matrici possono essere usate al posto dei dati grezzi per effettuare un'analisi univariata della varianza (con il sottocomando `MATRIX`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
23

Analisi GLM univariato

La procedura GLM univariato consente di eseguire un'analisi di regressione e un'analisi della varianza per una variabile dipendente tramite uno o più fattori e/o variabili. Le variabili fattore suddividono la popolazione in gruppi. Con questa procedura GLM (modello lineare generalizzato, General Linear Model) è possibile verificare ipotesi nulle relative agli effetti di altre variabili sulle medie di vari raggruppamenti di una sola variabile dipendente. È possibile analizzare le interazioni tra fattori e gli effetti di singoli fattori, alcuni dei quali possono essere casuali. È inoltre possibile includere gli effetti delle covariate e le interazioni tra covariate e fattori. Nell'analisi di regressione, le variabili indipendenti (stimatori) vengono specificate come covariate.

È possibile verificare sia modelli bilanciati che modelli non bilanciati. Un disegno è bilanciato se ciascuna cella del modello include lo stesso numero di casi. Oltre alla verifica delle ipotesi, la procedura GLM univariato consente di ottenere stime dei parametri.

Per la verifica di ipotesi sono disponibili contrasti a priori usati di frequente. Dopo che da un test *F* globale è risultata una certa significatività, è inoltre possibile eseguire test post hoc per valutare le differenze tra medie specifiche. L'opzione Medie marginali stimate consente di ottenere stime dei valori medi previsti delle celle incluse nel modello. I grafici di profilo, o grafici di interazione, di tali medie consentono di visualizzare in modo semplice alcune delle relazioni.

Residui, valori attesi, distanza di Cook e valori d'influenza possono essere salvati come variabili nel file di dati per la verifica di ipotesi.

Minimi quadrati ponderati consente di specificare una variabile per l'assegnazione di pesi diversi alle osservazioni per un'analisi di minimi quadrati ponderati (WLS), in alcuni casi per compensare la diversa precisione della misura.

Esempio. Per diversi anni vengono raccolti i dati relativi ai singoli partecipanti alla maratona di Chicago. Il tempo impiegato da ciascun partecipante per completare la maratona è la variabile dipendente. Altri fattori presi in considerazione sono le

condizioni metereologiche (freddo, caldo o temperatura moderata), il numero di mesi di allenamento, il numero di maratone corse in precedenza e il sesso. L'età è considerata una covariata. Dallo studio può risultare che il sesso rappresenta un effetto significativo, così come l'interazione tra sesso e condizioni meteorologiche.

Metodi. Per la valutazione di ipotesi diverse è possibile usare la somma dei quadrati Tipo I, Tipo II, Tipo III e Tipo IV. Il metodo di default è il Tipo III.

Statistiche. I seguenti test post hoc di intervalli e confronti multipli: Differenza meno significativa (LSD), Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch multiplo basato su test *F*, Ryan-Einot-Gabriel-Welsch a intervallo multiplo, Student-Newman-Keuls, differenze significative di Tukey, *b* di Tukey, Duncan, Hochberg (GT2), Gabriel, *t* di Waller-Duncan, Dunnett (a una e a due vie), Tamhane (T2), Dunnett (T3), Games-Howell e *C* di Dunnett. Statistiche descrittive: medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. Test di Levene di omogeneità della varianza.

Grafici. Grafici di variabilità vs. densità, dei residui e di profilo (interazione).

Dati. La variabile dipendente è quantitativa. I fattori sono categoriali. Vi possono essere associati valori numerici o valori stringa composti da un massimo di otto caratteri. Le covariate sono variabili quantitative correlate alla variabile dipendente.

Assunzioni. I dati sono costituiti da un campione casuale derivato da una popolazione normale in cui tutte le varianze di cella sono uguali. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. Per la verifica di ipotesi, è possibile usare test di omogeneità della varianza e i grafici di variabilità vs. intensità. È inoltre possibile esaminare residui e grafici dei residui.

Figura 23-1
Outuput di GLM univariato

Test degli Effetti tra soggetti

Variabile dipendente: SPVOL

Origine	Tipo III della Somma dei quadrati	df	Media dei quadrati	F	Sig.
Modello corretto	22.520[1]	11	2.047	12.376	.000
Intercetta	1016.981	1	1016.981	6147.938	.000
Farina	8.691	3	2.897	17.513	.000
Grasso	10.118	2	5.059	30.583	.000
Reagente	.997	2	.499	3.014	.082
Grasso* Reagente	5.639	4	1.410	8.522	.001
Errore	2.316	14	.165		
Totale	1112.960	26			
Totale corretto	24.835	25			

[1] R-quadrato = .907 (R-quadrato corretto = 833)

grasso* reagente

Variabile dipendente: SPVOL

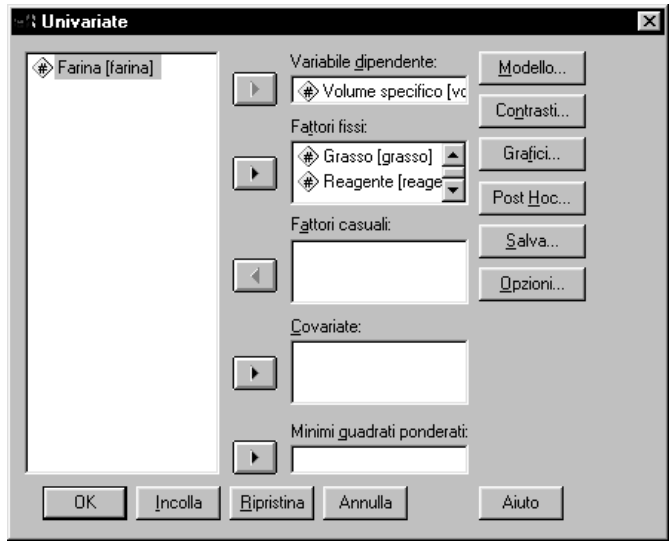
Grasso	Reagente	Media	Deviazione std.	Intervallo di confidenza al 95%	
				Limite inferiore	Limite superiore
1	1	5.536	.240	5.021	6.052
	2	5.891	.239	5.378	6.404
	3	6.123	.241	5.605	6.641
2	1	7.023	.241	6.505	7.541
	2	6.708	.301	6.064	7.353
	3	6.000	.203	5.564	6.436
3	1	6.629	.301	5.984	7.274
	2	7.200	.203	6.764	7.636
	3	8.589	.300	7.945	9.233

Per ottenere tabelle di GLM univariato

▶ Dai menu, scegliere:

Analizza
  Modello lineare generalizzato
    Univariata...

Figura 23-2
Finestra di dialogo Univariata



- Selezionare una variabile dipendente.
- Selezionare le variabili per l'opzione Fattori fissi, Fattori casuali o Covariate, a seconda dei dati in uso.
- È inoltre possibile utilizzare l'opzione Minimi quadrati ponderati per specificare l'analisi dei minimi quadrati ponderati. Casi con valore 0, negativo o mancante per la variabile di ponderazione saranno esclusi dall'analisi. Non è possibile utilizzare come variabile di ponderazione una variabile già inclusa nel modello.

GLM – Univariato: Modello

Figura 23-3
Finestra di dialogo Univariata: Modello

Specifica modello. Un modello fattoriale completo contiene tutti gli effetti principali dei fattori e delle covariate e tutte le interazioni fattore per fattore. Non contiene interazioni di covariate Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

Fattori e covariate. Visualizza un elenco dei fattori e delle covariate in cui (F) indica un fattore fisso e (C) una covariata. In un'analisi univariata (R) indica un fattore casuale.

Modello. Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire.

Somma dei quadrati. Metodo per il calcolo della somma dei quadrati. Il metodo Somma dei quadrati in genere utilizzato con modelli bilanciati o non bilanciati privi di celle mancanti è il Tipo III.

Includi l'intercetta nel modello. L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

Costruisci termini

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione di default.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Somma dei quadrati

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo di default, è quello usato più di frequente.

Tipo I. Questo metodo è definito anche scomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene corretto solo per i termini del modello che lo precedono. Il metodo Somma dei quadrati Tipo I è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

Tipo II. Questo metodo consente di calcolare le somme dei quadrati di un effetto del modello corretto per tutti gli altri effetti "appropriati". È considerato appropriato un effetto corrispondente a tutti gli effetti che non includono l'effetto in esame. Il metodo Somma dei quadrati Tipo II è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato.
- Qualsiasi modello che include solo effetti principali del fattore.
- Qualsiasi modello di regressione.
- Un disegno nidificato in modo puro. Questo tipo di nidificazione può essere specificato tramite la sintassi.

Tipo III. Tipo di default. Questo metodo consente di calcolare le somma dei quadrati di un effetto del disegno come la somma dei quadrati corretta per qualsiasi altro effetto che non lo include e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie ponderate di Yates. Il metodo Somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

Tipo IV. Questo metodo è specifico per situazioni con celle mancanti. Per qualsiasi effetto *F* del disegno, se *F* non è incluso in nessun altro effetto, allora Tipo IV = Tipo III = Tipo II. Se invece *F* è incluso in altri effetti, con il Tipo IV i contrasti creati tra i parametri in *F* vengono distribuiti equamente tra tutti gli effetti di livello superiore. Il metodo Somma dei quadrati Tipo IV viene in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato e non bilanciato contenente celle vuote.

GLM - Univariato: Contrasti

Figura 23-4
Finestra di dialogo Univariata: Contrasti

I contrasti consentono di verificare il grado di differenza tra i livelli di un fattore. È possibile specificare un contrasto per ciascun fattore del modello (in un modello a misure ripetute, un contrasto per ciascun fattore tra soggetti). I contrasti rappresentano combinazioni lineari dei parametri.

La verifica di ipotesi è basata sull'ipotesi nulla **LB** = 0, dove **L** è la matrice dei coefficienti di contrasto e **B** è il vettore dei parametri. Quando viene specificato un contrasto, SPSS crea una matrice **L** in cui le colonne relative al fattore corrispondono al contrasto. Le altre colonne vengono corrette in modo che la matrice **L** possa essere stimata.

L'output include una statistica *F* per ciascun insieme di contrasti. Per le differenze dei contrasti vengono inoltre visualizzati gli intervalli di confidenza simultanei di tipo Bonferroni basati su una distribuzione *t* di Student.

Contrasti disponibili

Sono disponibili i contrasti deviazione, semplici, differenza, Helmert, ripetuti e polinomiali. Per i contrasti deviazione e i contrasti semplici, è possibile stabilire se la categoria di riferimento corrisponde alla prima o all'ultima categoria.

Tipi di contrasto

Deviazione. Consente di confrontare la media di ciascun livello, a eccezione di una categoria di riferimento, con la media di tutti i livelli (media principale). L'ordine dei livelli del fattore può essere un ordine qualsiasi.

Semplice. Consente di confrontare la media di ciascun livello con la media di un livello specifico. Questo tipo di contrasto risulta utile quando è disponibile un gruppo di controllo. Come categoria di riferimento, è possibile scegliere la prima o l'ultima categoria.

Differenza. Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Questo tipo di contrasto è a volte definito contrasto inverso di Helmert.

Helmert. Consente di confrontare la media di ciascun livello del fattore (a eccezione dell'ultimo) con la media dei livelli successivi.

Ripetuto. Consente di confrontare la media di ciascun livello (a eccezione dell'ultimo) con la media del livello successivo.

Polinomiale. Consente di confrontare l'effetto lineare, quadratico, cubico e così via. Tutte le categorie del primo grado di libertà includono l'effetto lineare, quelle del secondo includono l'effetto quadratico e così via. Questi contrasti sono spesso usati per la stima di trend polinomiali.

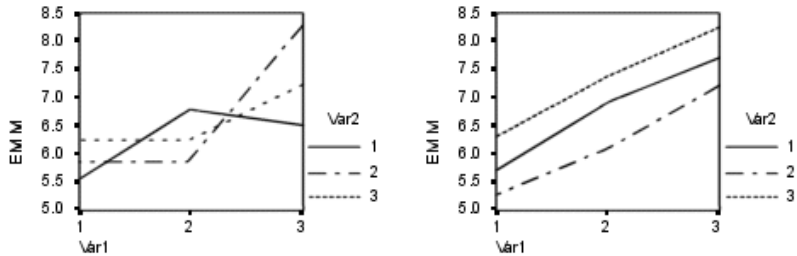
GLM – Univariato: Profili

Figura 23-5
Finestra di dialogo Univariata: Profili

I profili, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un profilo è un grafico lineare in cui ciascun punto indica la media marginale stimata di una variabile dipendente (corretta per le covariate) in corrispondenza di un solo livello di un fattore. È possibile utilizzare i livelli di un secondo fattore per creare linee distinte. È possibile utilizzare ciascun livello di un terzo fattore per creare un grafico distinto. Tutti gli eventuali fattori casuali e fissi sono disponibili per i grafici. In analisi multivariate i grafici di profilo vengono creati per ciascuna variabile dipendente. Nei grafici di profilo per un'analisi a misure ripetute è possibile includere sia fattori tra soggetti che fattori entro soggetti. GLM multivariato e a misure ripetute sono disponibili solo se è stata installata l'opzione Advanced Models.

Il profilo di un fattore mostra se le medie marginali stimate aumentano o diminuiscono tra i vari livelli. Nel caso di due o più fattori, le linee parallele indicano che tra i fattori non esiste alcuna interazione, ovvero che è possibile analizzare i livelli di un solo fattore. Le linee che si incrociano indicano invece che esiste un'interazione.

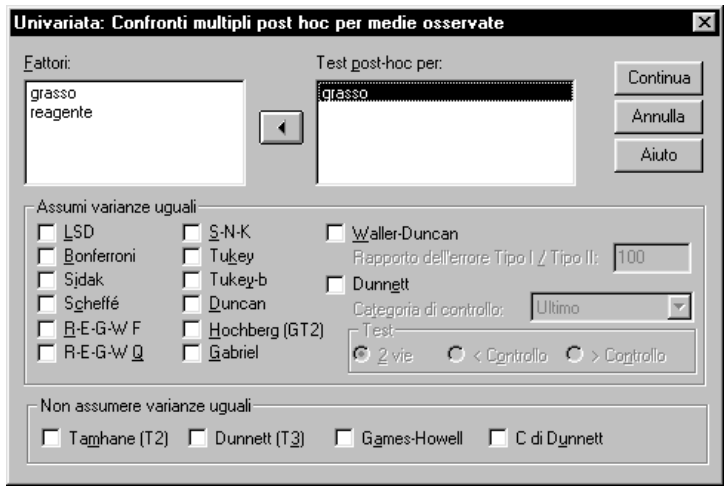
Figura 23-6
Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)



I grafici definiti tramite la selezione dei fattori per l'asse orizzontale e, se lo si desidera, dei fattori di linee e di grafici separati devono essere inclusi nell'elenco dei grafici.

GLM - Univariato: Confronti post hoc

Figura 23-7
Finestra di dialogo Univariata: Confronti multipli post hoc per medie osservate



Test per confronti multipli post hoc. Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervalli e i confronti a coppie multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stata apportata alcuna correzione. Questi test vengono usati solo con fattori fissi tra soggetti. Nella procedura GLM a misure ripetute, questi test non sono disponibili se non sono presenti fattori tra soggetti e i test per confronti multipli post hoc vengono eseguiti per la media tra i livelli dei fattori entro soggetti. Per la procedura GLM multivariato, i test post hoc vengono eseguiti separatamente per ciascuna variabile dipendente. GLM multivariato e a misure ripetute sono disponibili solo se è stata installata l'opzione Advanced Models.

I test per confronti multipli usati più di frequente sono il test di Bonferroni e il test delle differenze significative di Tukey. Il **test di Bonferroni**, basato sulla statistica *t* di Student, consente di correggere il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test t di Sidak** corregge inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test delle differenze significative di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti a coppie tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti per coppie. Quando si eseguono test su un elevato numero di coppie di medie, il test delle differenze significative

di Tukey risulta più efficace rispetto al test di Bonferroni. Nel caso di un numero limitato di coppie, risulta invece più efficace il test di Bonferroni.

Il test di **Hochberg (GT2)** è simile al test delle differenze significative di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test dei confronti a coppie di Gabriel** utilizza il modulo massimo studentizzato ed è in genere più indicativo del test di Hochberg (GT2) quando le dimensioni delle celle sono diverse. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.

Il **test t per confronti multipli a coppie di Dunnett** confronta un insieme di trattamenti con una media di controllo singola. L'ultima categoria è la categoria di controllo di default. In alternativa, è possibile scegliere la prima categoria. È inoltre possibile scegliere un test a 2 vie oppure a 1 via. Per verificare che la media in qualsiasi livello del fattore (a eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo, è necessario usare un test a due sensi. Per verificare se la media di qualsiasi livello del fattore è minore di quella della categoria di controllo, selezionare < Controllo. In modo analogo, per verificare se la media di qualsiasi livello del fattore è maggiore di quella della categoria di controllo, selezionare > Controllo.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test a intervalli decrescenti multipli. Le procedure a multipli decrescenti verificano in primo luogo se tutte le medie sono uguali. Se le medie non risultano tutte uguali, il test di uguaglianza viene eseguito su un sottoinsieme di medie. Il test **R-E-G-W F** è basato su un test *F*, mentre **R-E-G-W Q** è basato sull'intervallo studentizzato. Questi test risultano più efficaci rispetto ai test a intervallo multiplo di Duncan e Student-Newman-Keuls, che sono pure procedure a intervalli decrescenti multipli. È tuttavia consigliabile non usarli con celle di dimensioni non uguali.

Quando le varianze non sono uguali, è necessario utilizzare il test **Tamhane (T2)** (test per confronti a coppie conservativo basato su un test*t*), il test di **Dunnett T3** (test per confronti a coppie basato sul modulo studentizzato), il test per confronti a coppie di **Games-Howell** (a volte poco conservativo) o il test **C di Dunnett**(test per confronti a coppie basato sull'intervallo studentizzato).

Il test a intervallo multiplo di Duncan, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test per intervallo che classificano le medie raggruppate e calcolano un valore di intervallo. Questi test sono usati meno frequentemente dei test descritti in precedenza.

Il **test t di Waller-Duncan** utilizza un approccio bayesiano. Si tratta di un test a intervallo che usa la media armonica della dimensione campione nel caso di dimensioni campione non uguali.

Il livello di significatività del **test di Scheffé** consente la verifica di tutte le possibili combinazioni lineari delle medie di gruppo e non dei soli confronti a coppie disponibili in questa funzione. Il test di Scheffé risulta pertanto più conservativo rispetto ad altri test, ovvero per ottenere un livello sufficiente di significatività, è richiesta una differenza maggiore tra le medie.

Il test per confronti a coppie multipli Differenza meno significativa o **LSD**, è equivalente a più test *t* tra tutte le coppie di gruppi. Lo svantaggio di questo test è che non viene eseguito alcun tentativo di correzione del livello di significatività osservata per confronti multipli.

Test visualizzati. I confronti a coppie sono disponibili per i test LSD, Sidak, Bonferroni, Games e Howell, Tamhane (T2) e (T3), test *C* di Dunnett e Dunnett (T3). Per i test S-N-K, *b* di Tukey, Duncan, R-E-G-W *F*, R-E-G-W *Q* e Waller sono disponibili sottoinsiemi omogenei per test per intervallo. Il test delle differenze significative di Tukey, i test Hochberg (GT2), Gabriel e Scheffé sono sia test per confronti multipli che test a intervallo.

GLM – Univariato: Salva

Figura 23-8
Finestra di dialogo Univariata: Salva

È possibile salvare i valori attesi dal modello, le misure correlate e i residui come nuove variabili nell'Editor dei dati. Molte di queste variabili possono essere usate per l'esame di ipotesi sui dati. Per salvare i valori in modo da poterli usare in un'altra sessione SPSS, è necessario salvare il file di dati corrente.

Valori attesi. Valori attesi dal modello per ciascun caso.

- **Non standardizzati.** I valori risultanti dal modello per la variabile dipendente e per ciascun caso.
- **Pesati.** I valori attesi pesati non standardizzati. Disponibile solo se è stata selezionata una variabile WLS.
- **Errore standard.** Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

Diagnostici. Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere una notevole influenza sul modello.

- **distanza di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione cambierebbe sostanzialmente il loro valore.
- **Valori d'influenza.** L'influenza relativa di ogni osservazione sull'adattamento del modello.

Residui. Un residuo non standardizzato corrisponde al valore effettivo della variabile dipendente diminuito del valore atteso dal modello. Sono inoltre disponibili residui standardizzati, studentizzati ed eliminati. Se è stata selezionata una variabile WLS, saranno inoltre disponibili residui non standardizzati pesati.

- **Non standardizzati.** La differenza tra un valore osservato e il valore stimato dal modello.
- **Pesati.** I residui pesati non standardizzati. Disponibile solo se è stata selezionata una variabile WLS.
- **Standardizzati.** Il residuo standardizzato, conosciuto anche come residuo di Pearson, è il residuo semplice diviso per l'errore standard della frequenza osservata. Benché i residui standardizzati siano asintoticamente normali, si comportano in modo meno variabile (in termini di varianza) delle variabili casuali normali. La loro distribuzione ha media 0 e deviazione standard 1

- **Studentizzati.** Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, in funzione della distanza tra ogni valore delle variabili indipendenti per un caso e la media di tutte le variabili indipendenti per lo stesso caso. Il residuo studentizzato riflette in maniera più precisa rispetto al residuo standardizzato le variazioni dell'errore da caso a caso.
- **Con rimozione.** Il residuo per un caso se quel caso venisse escluso dal calcolo dei coefficienti di regressione. È la differenza fra il valore della variabile dipendente e l valore stimato corretto.

Statistiche coefficienti. Scrive una matrice della varianza-covarianza delle stime dei parametri nel modello in un nuovo file di dati della sessione attiva o in un file di dati esterno in formato SPSS. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime dei parametri, una riga di valori di significatività per le statistiche *t* corrispondenti alle stime dei parametri e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente di modelli multivariati sono disponibili righe simili. Questi dati possono essere usati in altre procedure SPSS che eseguono la lettura di un file di matrici SPSS.

GLM - Univariato: Opzioni

Figura 23-9
Finestra di dialogo Univariata: Opzioni

In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

Medie marginali stimate. Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono corrette per le eventuali covariate.

- **Confronta effetti principali.** Consente di eseguire confronti a coppie senza correzione tra le medie marginali stimate di qualsiasi effetto principale del modello, per fattori sia tra soggetti che entro soggetti. Questa opzione è disponibile solo se nell'elenco Medie marginali per sono stati selezionati effetti principali.
- **Correzione intervallo di confidenza.** Selezionare la differenza meno significativa (LSD), la correzione di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo comando è disponibile solo se è stato selezionato Confronta effetti principali.

Visualizzazione. Selezionare Statistiche descrittive per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione Stima della dimensione degli effetti fornisce un valore eta-quadrato parziale per ciascun effetto e per ciascuna stima dei parametri. La statistica eta-quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare Potenza osservata per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare Stime dei parametri per ottenere stime dei parametri, errori standard, test *t*, intervalli di confidenza e la potenza osservata per ciascun test. Selezionare Matrice dei coefficienti di contrasto per ottenere la matrice **L**.

La funzione Test di omogeneità produce il test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori fra soggetti, solo per i fattori fra soggetti. Le opzioni Grafici di variabilità vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare Grafici dei residui per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare Mancanza di adattamento per controllare che la relazione fra la variabile dipendente e le variabili indipendenti possa essere descritta in modo adeguato dal modello. La funzione Forma funzionale generalizzata consente di creare test di ipotesi personalizzati basati sulla forma funzionale generalizzata. Le righe di una matrice dei coefficienti di contrasto sono combinazioni lineari della forma funzionale generalizzata.

Livello di significatività. Potrebbe risultare utile correggere il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

Funzioni aggiuntive del comando UNIANOVA

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Specificare gli effetti nidificati del disegno (tramite il sottocomando `DESIGN`).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `TEST`).
- Specificare contrasti multipli (tramite il sottocomando `CONTRAST`).

- Includere valori mancanti definiti dall'utente (tramite il sottocomando MISSING).
- Specificare criteri EPS (tramite il sottocomando CRITERIA).
- Costruire una matrice **L**, **M** o **K** personalizzata (tramite il sottocomando LMATRIX, MMATRIX o KMATRIX).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando CONTRAST).
- Specificare metrica per contrasti polinomiali (tramite il sottocomando CONTRAST).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando POSTHOC).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori per i fattori elencati (tramite il sottocomando EMMEANS).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando SAVE).
- Costruire un file di dati di matrici di correlazione (tramite il sottocomando OUTFILE).
- Costruire un file di dati di matrici contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando OUTFILE).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando OUTFILE).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
24

Correlazioni bivariate

La procedura Correlazioni bivariate consente di calcolare coefficiente di correlazione di Pearson, rho di Spearman e tau-*b* di Kendall con i rispettivi livelli di significatività. Le correlazioni consentono di misurare la relazione tra variabili o punteggi di rango. Prima di calcolare un coefficiente di correlazione, è necessario valutare la presenza di valori anomali nei dati (che possono causare risultati errati) e l'esistenza di una relazione lineare. Il coefficiente di correlazione di Pearson è una misura di associazione lineare. Due variabili possono essere perfettamente correlate, ma se la relazione non è lineare il coefficiente di correlazione di Pearson non è la statistica migliore per misurare tale associazione.

Esempio. Il numero di partite vinte da una squadra di baseball è correlato con la media dei punti totalizzati per ciascuna partita? Un grafico a dispersione indica l'esistenza di una relazione lineare. Dall'analisi dei dati relativi alla stagione NBA 1994–1995 risulta che il coefficiente di correlazione di Pearson (0,581) è significativo al livello 0,01. Si può presumere che il numero di partite vinte per stagione sia inversamente proporzionale ai punti totalizzati dagli avversari. Queste variabili sono legate da una correlazione negativa (–0,401), significativa al livello 0,05.

Statistiche. Per ogni variabile: numero di casi con valori non mancanti, media e deviazione standard. Per ciascuna coppia di variabili: coefficiente di correlazione di Pearson, rho di Spearman, tau-*b* di Kendall, prodotto incrociato delle deviazioni, covarianza.

Dati. Utilizzare le variabili quantitative simmetriche per il coefficiente di correlazione di Pearson e le variabili quantitative o le variabili con categorie ordinate per rho di Spearman e tau-*b* di Kendall.

Assunzioni. Il coefficiente di correlazione di Pearson assume che ciascuna coppia di variabili sia bivariata normale.

Figura 24-1
Output di correlazioni bivariate

Correlazioni

		Partite vinte	Punti realizzati per partita	Punti subiti per partita
Correlazione di Pearson	Partite vinte	1.000	.581**	-.401*
	Punti realizzati per partita	.581**	1.000	.457*
	Punti subiti per partita	-.401*	.457*	1.000
Sig. (2-code)	Partite vinte	.	.001	.038
	Punti realizzati per partita	.001	.	.017
	Punti subiti per partita	.038	.017	.
N	Partite vinte	27	27	27
	Punti realizzati per partita	27	27	27
	Punti subiti per partita	27	27	27

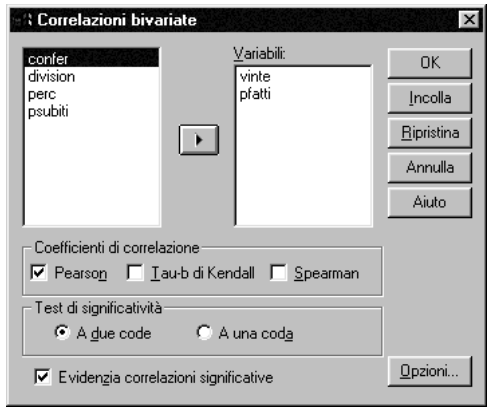
**. La correlazione è significativa al livello 0,01 (2-code).
*. La correlazione è significativa al livello 0,05 (2-code).

Per ottenere correlazioni bivariate

Dai menu, scegliere:

Analizza
  Correlazione
    Bivariata...

Figura 24-2
Finestra di dialogo Correlazioni bivariate



▶ Selezionare una o più variabili numeriche.

Sono inoltre disponibili le seguenti opzioni:

- **Coefficienti di correlazione.** Per variabili quantitative normalmente distribuite, scegliere il coefficiente di correlazione di Pearson. Se i dati non sono normalmente distribuiti o prevedono categorie ordinate, scegliere Tau-b di Kendall o Spearman, per misurare l'associazione tra punteggi di rango. I valori dei coefficienti di correlazione variano da –1 (relazione negativa perfetta) a +1 (relazione positiva perfetta). Il valore 0 indica l'assenza di relazione lineare. Interpretando i risultati, evitare di trarre conclusioni di tipo causa-effetto sulla base di una correlazione significativa.
- **Test di significatività.** È possibile selezionare le probabilità a una coda o a due code. Se si conosce in anticipo la direzione dell'associazione, selezionare A una coda.. In alternativa, selezionare A due code.
- **Evidenzia correlazioni significative.** I coefficienti di correlazione significativi al livello 0,05 vengono identificati con un asterisco singolo e quelli significativi al livello 0,01 con due asterischi.

Correlazioni bivariate: Opzioni

Figura 24-3
Finestra di dialogo Correlazioni bivariate: Opzioni

Statistiche. Per le correlazioni di Pearson, è possibile scegliere una o entrambe le seguenti opzioni.

- **Medie e deviazioni standard.** Visualizzate per ciascuna variabile. Viene indicato anche il numero di casi con valori non mancanti. I valori mancanti vengono gestiti variabile per variabile indipendentemente dall'impostazione corrispondente.
- **Prodotti degli scarti e delle covarianze.** Viene visualizzato per ciascuna coppia di variabili. La deviazione del prodotto incrociato è equivalente alla somma dei prodotti delle variabili corrette per la media. È il numeratore del coefficiente di correlazione di Pearson. La covarianza è una misura non standardizzata della relazione tra due variabili, equivalente alla deviazione del prodotto incrociato divisa per *N*–1.

Valori mancanti. È possibile scegliere tra le opzioni seguenti:

- **Esclusione pairwise.** I casi con valori mancanti per una o entrambe le variabili di una coppia per un coefficiente di correlazione vengono esclusi dall'analisi. Poiché ciascun coefficiente si basa su tutti i casi con codici validi per quella particolare coppia di variabili, in tutti i calcoli verrà utilizzato il maggior numero di informazioni disponibile. In questo modo è possibile ottenere una serie di coefficienti basati su un numero variabile di casi.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile vengono esclusi da tutte le correlazioni.

Funzioni aggiuntive dei comandi CORRELATIONS e NONPAR CORR

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Scrivere una matrice di correlazione per la correlazione di Pearson da utilizzare in luogo dei dati per ottenere altri tipi di analisi, ad esempio l'analisi fattoriale (con il sottocomando `MATRIX`).
- Ottenere correlazioni di ciascuna variabile presente in un elenco con le variabili corrispondenti presenti in un secondo elenco (utilizzando la parola chiave `WITH` con il sottocomando `VARIABLES`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
25

Correlazioni parziali

La procedura Correlazioni parziali consente di calcolare i coefficienti di correlazione parziale che descrivono la relazione lineare tra due variabili controllando gli effetti di una o più variabili aggiuntive. Le correlazioni sono misure di associazione lineare. Due variabili possono essere perfettamente correlate, ma se la relazione non è lineare, un coefficiente di correlazione non rappresenta la statistica più adatta a misurarne l'associazione.

Esempio. Esiste una correlazione tra i fondi stanziati per il sistema sanitario e il tasso di malattie? Sebbene si potrebbe pensare che una relazione di tale tipo sia negativa, da uno studio emerge una correlazione *positiva* significativa: con l'aumentare dei fondi stanziati per il sistema sanitario, il tasso di malattie sembra aumentare. Il controllo della frequenza delle visite ai medici elimina virtualmente la correlazione positiva osservata. I fondi stanziati per il sistema sanitario e il tasso di malattie sembrano avere una correlazione positiva solo perché più pazienti hanno accesso al sistema sanitario quando vengono aumentati i fondi, con un conseguente aumento delle malattie segnalate da medici e ospedali.

Statistiche. Per ogni variabile: numero di casi con valori non mancanti, media e deviazione standard. Matrici di correlazione parziale e di ordine zero, con gradi di libertà e livelli di significatività.

Dati. Utilizzare variabili simmetriche, quantitative.

Assunzioni. La procedura Correlazioni parziali si fonda sull'ipotesi che ciascuna coppia di variabili sia bivariata normale.

Figura 25-1
Output delle correlazioni parziali

Correlazioni

Variabili di controllo			Health care funding (amount per 100)	Reported diseases (rate per 10,000)	Visits to health care providers (rate per 10,000)
-nessuno-[a]	Health care funding (amount per 100)	Correlazione	1,000	,737	,964
		Significatività (a due code)	.	,000	,000
		df	0	48	48
	Reported diseases (rate per 10,000)	Correlazione	,737	1,000	,762
		Significatività (a due code)	,000	.	,000
		df	48	0	48
	Visits to health care providers (rate per 10,000)	Correlazione	,964	,762	1,000
		Significatività (a due code)	,000	,000	.
		df	48	48	0
Visits to health care providers (rate per 10,000)	Health care funding (amount per 100)	Correlazione	1,000	,013	
		Significatività (a due code)	.	,928	
		df	0	47	
	Reported diseases (rate per 10,000)	Correlazione	,013	1,000	
		Significatività (a due code)	,928	.	
		df	47	0	

a. Le celle contengono correlazioni di ordine zero (Pearson).

Per ottenere correlazioni parziali

► Dai menu, scegliere:

Analizza
  Correlazione
    Parziale...

Figura 25-2
Finestra di dialogo Correlazioni parziali

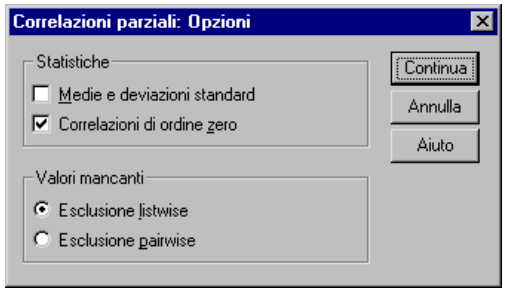
- Selezionare due o più variabili numeriche per cui è necessario calcolare le correlazioni parziali.
- Selezionare una o più variabili numeriche di controllo.

Sono inoltre disponibili le seguenti opzioni:

- **Test di significatività.** È possibile selezionare le probabilità a una coda o a due code. Se si conosce in anticipo la direzione dell'associazione, selezionare A una coda.. In alternativa, selezionare A due code.
- **Mostra l'esatto livello di significatività.** Per impostazione predefinita, vengono indicati la probabilità e i gradi di libertà di ciascun coefficiente di correlazione. Se questa opzione viene deselezionata, i coefficienti significativi al livello 0,05 vengono identificati con un solo asterisco, i coefficienti significativi al livello 0,01 con un doppio asterisco e i gradi di libertà vengono eliminati. Questa impostazione viene applicata alle matrici di correlazione parziale e di ordine zero.

Correlazioni parziali: Opzioni

Figura 25-3
Finestra di dialogo Correlazioni parziali: Opzioni



Statistiche. È possibile scegliere una delle seguenti opzioni o entrambe:

- **Medie e deviazioni standard.** Visualizzate per ciascuna variabile. Viene indicato anche il numero di casi con valori non mancanti.
- **Correlazioni di ordine zero.** Viene visualizzata una matrice di correlazioni semplici tra tutte le variabili, incluse le variabili di controllo.

Valori mancanti. È possibile scegliere una delle seguenti opzioni:

- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile, incluse le variabili di controllo, vengono esclusi da tutti i calcoli.
- **Esclusione pairwise.** Per il calcolo delle correlazioni di ordine zero su cui si basano le correlazioni parziali, i casi con valori mancanti per una o entrambe le variabili di una coppia non verranno utilizzati. L'eliminazione pairwise consente di utilizzare il massimo numero di dati possibile. Il numero di casi, tuttavia, può differire a seconda del coefficiente. Quando è attiva l'eliminazione pairwise, i gradi di libertà per un particolare coefficiente parziale si basano sul numero minimo di casi utilizzati per il calcolo di una delle correlazioni di ordine zero.

Opzioni aggiuntive del comando PARTIAL CORR

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Leggere una matrice di correlazione di ordine zero o scrivere una matrice di correlazione parziale (con il sottocomando `MATRIX`).
- Ottenere correlazioni parziali tra due elenchi di variabili (con la parola chiave `WITH` nel sottocomando `VARIABLES`).

- Ottenere più analisi (con più sottocomandi `VARIABLES`).
- Specificare i valori degli ordini da richiedere (ad esempio sia le correlazioni parziali di primo e secondo ordine) quando sono disponibili due variabili di controllo (con il sottocomando `VARIABLES`).
- Eliminare i coefficienti ridondanti (con il sottocomando `FORMAT`).
- Visualizzare una matrice di correlazioni semplici quando non è possibile calcolare alcuni coefficienti (con il sottocomando `STATISTICS`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
26

Distanze

Questa procedura consente di calcolare una grande varietà di statistiche in base alle similarità o alle dissimilarità (distanze), sia considerando coppie di variabili, sia considerando coppie di casi. Tali misure di distanza o similarità possono poi essere utilizzate insieme ad altre procedure quali analisi fattoriale, analisi cluster o scaling multidimensionale per l'analisi di insiemi complessi di dati.

Esempio. È possibile misurare le similarità tra coppie di automobili in base a caratteristiche specifiche, quali dimensione del motore, MPG e potenza. Grazie al calcolo delle similarità, è possibile stabilire quali automobili sono simili e quali sono differenti tra loro. Per un'analisi più formale, si può scegliere di applicare alle similarità l'analisi cluster gerarchica o lo scaling multidimensionale che consentono di esplorare la struttura sottostante.

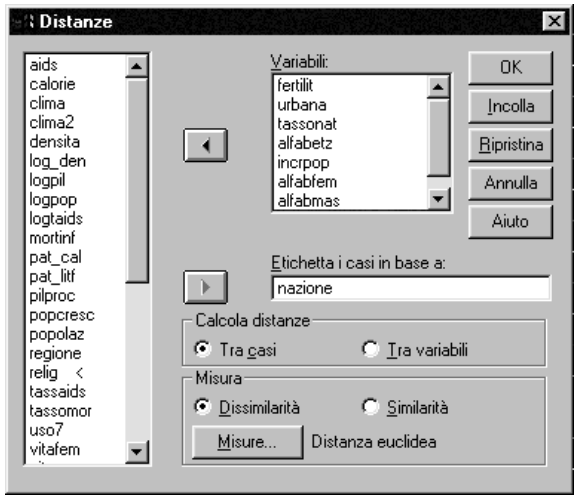
Statistiche. Le misure di dissimilarità (distanza) disponibili per i dati di intervallo sono: distanza euclidea, distanza euclidea quadratica, Chebychev, City-block, Minkowski, personalizzata. Per i dati di conteggio, le misure disponibili sono chi-quadrato o phi-quadrato. Per i dati binari, infine, le misure disponibili sono: distanza euclidea, distanza euclidea quadratica, differenza di dimensione, differenza di modello, varianza, forma, Lance e Williams. Le misure di similarità disponibili per i dati di intervallo sono la correlazione di Pearson o il coseno. Per i dati binari sono invece disponibili le seguenti misure: Russel e Rao, corrispondenza semplice, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* di Anderberg, *Y* di Yule, *Q* di Yule, Ochiai, Sokal e Sneath 5, correlazione phi a 4 punti o dispersione.

Per ottenere matrici delle distanze

- Dai menu, scegliere:

  Analizza
  Correlazione
  Distanze...

Figura 26-1
Finestra di dialogo Distanze

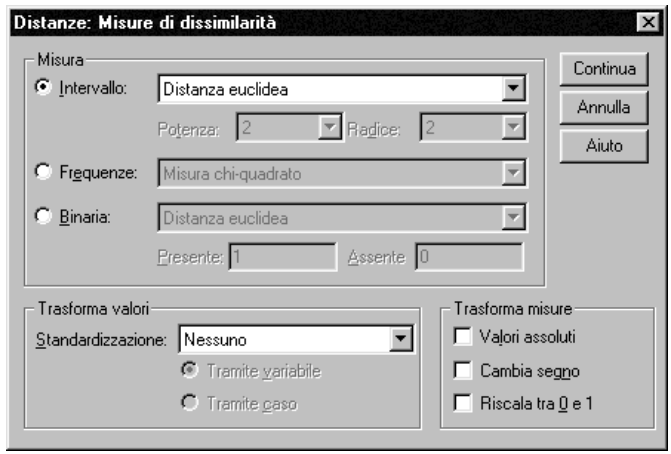


- Selezionare una variabile numerica per calcolare la distanza tra casi e almeno due variabili numeriche per calcolare la distanza tra variabili.

- Scegliere un'alternativa nel gruppo Calcola distanze per calcolare le distanze tra casi o tra variabili.

Distanze: Misure di dissimilarità

Figura 26-2
Finestra di dialogo Distanze: Misure di dissimilarità



Dal gruppo Misura selezionare l'alternativa corrispondente al tipo di dati desiderato (intervallo, conteggio o binari), quindi scegliere dall'elenco a discesa una delle misure corrispondenti a quel tipo di dati. Le misure disponibili per tipo di dati sono le seguenti:

- **Dati per intervallo.** Distanza euclidea, distanza euclidea quadratica, Chebychev, City-block, Minkowski o personalizzata.
- **Conteggio dati.** Misura chi-quadrato e misura phi-quadrato.
- **Dati binari.** Distanza euclidea, distanza euclidea quadratica, differenza di dimensione, differenza di modello, varianza, forma o di Lance e Williams. (Inserire i valori Presente e Assente per specificare i due valori significativi, tutti gli altri valori verranno ignorati dalle distanze).

Il gruppo Trasforma valori consente di standardizzare i valori per casi o valori *prima* di calcolare le similarità. Tali trasformazioni non sono applicabili ai dati binari. I metodi di standardizzazione disponibili sono: punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, ampiezza massima di 1, media di 1 o deviazione standard di 1.

Il gruppo Trasforma misure consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura di distanza. Le alternative possibili sono: Valori assoluti, Cambia segno e Riscala all'intervallo 0–1.

Distanze: Misure di similarità

Figura 26-3
Finestra di dialogo Distanze: Misure di similarità

Selezionare l'alternativa corrispondente al tipo di dati desiderato (intervallo o binari) dal gruppo Misura , quindi scegliere una delle misure corrispondenti a quel tipo di dati dall'elenco a discesa. Le misure disponibili per tipo di dati sono le seguenti:

- **Dati per intervallo.** Correlazione Pearson o coseno.
- **Dati binari.** Russel e Rao, corrispondenza semplice, Jaccard, Dice, Rogers e Tanimoto, Sokal e Sneath 1, Sokal e Sneath 2, Sokal e Sneath 3, Kulczynski 1, Kulczynski 2, Sokal e Sneath 4, Hamann, Lambda, *D* di Anderberg, *Y* di Yule, *Q* di Yule, Ochiai, Sokal e Sneath 5, correlazione phi a 4 punti o dispersione. (Inserire i valori Presente e Assente per specificare i due valori significativi, tutti gli altri valori verranno ignorati dalle distanze).

Il gruppo Trasforma valori consente di standardizzare i valori per casi o valori prima di calcolare le similarità. Tali trasformazioni non sono applicabili ai dati binari. I metodi di standardizzazione disponibili sono: punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, ampiezza massima di 1, media di 1 e deviazione standard di 1.

Il gruppo Trasforma misure consente di trasformare i valori generati dalla misura della distanza. Questi verranno applicati dopo il calcolo della misura di distanza. Le alternative possibili sono: Valori assoluti, Cambia segno e Riscala all'intervallo 0–1.

Opzioni aggiuntive del comando PROXIMITIES

La procedura Distanze usa la sintassi del comando `PROXIMITIES`. Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Specificare un numero intero come potenza per la misura della distanza di Minkowski.
- Specificare qualsiasi intero come potenza e radice per la misura personalizzata delle distanze.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
27

Regressione lineare

La regressione lineare consente di stimare i coefficienti dell'equazione lineare, includendo una o più variabili indipendenti, che prevedono al meglio il valore della variabile dipendente. Ad esempio, è possibile tentare di prevedere le vendite annuali di un rappresentante (la variabile dipendente) in base a variabili indipendenti quali l'età, gli studi e gli anni di esperienza lavorativa.

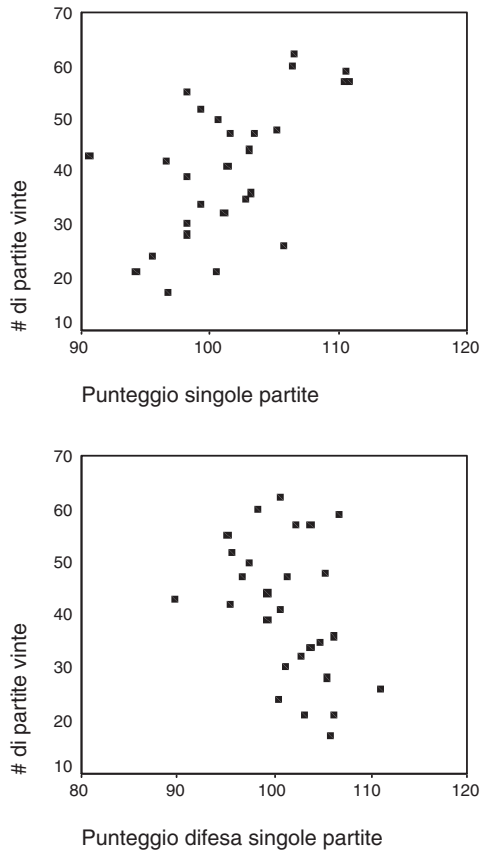
Esempio. Il numero di partite vinte da una squadra di basket in una stagione è correlato al numero medio di punti effettuati dalla squadra per partita? Un grafico a dispersione indica che queste variabili sono correlate in modo lineare. Il numero di partite vinte e il numero medio di punti effettuati dalla squadra avversaria sono anch'essi correlati in modo lineare. Queste variabili hanno una relazione negativa. Al crescere del numero delle partite vinte, diminuisce il numero medio di punti effettuati dall'avversario. Con la regressione lineare, è possibile modellare la relazione di queste variabili. È possibile utilizzare un modello valido per stimare quante partite vinceranno le squadre.

Statistiche. Per ogni variabile: numero di casi validi, media e deviazione standard. Per ogni modello: coefficienti di regressione, matrice di correlazione, correlazioni di ordine zero e parziali, *R* multipli, $R^{quadrato}$, R^2 corretto, variazioni $R^{quadrato}$, errore standard della stima, tabella di analisi della varianza, valori attesi e residui. Inoltre, intervalli di confidenza al 95% per ogni coefficiente di regressione, matrice di varianza-covarianza, fattore d'inflazione della varianza, tolleranza, test di Durbin-Watson, misure di distanza (Mahalanobis, Cook e valori di influenza), DiffBeta, DiffAdatt, intervalli di stima e diagnostiche per casi. Grafici: grafici a dispersione, grafici parziali, istogrammi e grafici di probabilità normale.

Dati. Le variabili dipendenti ed indipendenti devono essere quantitative. È necessario che le variabili categoriali, come la religione, l'età o la regione di residenza, siano ricodificate come variabili binarie (fittizie) o altri tipi di variabili di contrasto.

Assunzioni. Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. La relazione tra la variabile dipendente e ogni variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti.

Figura 27-1
Output di Regressione lineare



Riepilogo del modello [2]

Modello	R	R-quadrato	R-quadrato corretto	Errore std. della stima
1	.947[1]	.898	.889	4.40

1. Stimatori: (Costante), Punti subiti per partita, Punti realizzati per partita
2. Variabile dipendente: Partite vinte

ANOVA [2]

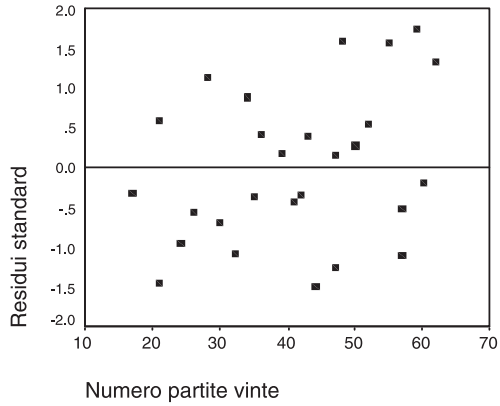
Modello		Somma dei quadrati	df	Media dei quadrati	F	Sig.
1	Regressione	4080.533	2	2040.266	105.198	.000[1]
	Residuo	465.467	24	19.394		
	Totale	4546.000	26			

1. Stimatori: (Costante), Punti subiti per partita, Punti realizzati per partita
2. Variabile dipendente: Partite vinte

Coefficienti[1]

Modello		Coefficienti non standardizzati		Coefficienti standardizzati	t	Sig.
		B	Errore std.	Beta		
1	(Costante)	28.121	21.404		1.314	.201
	Punti realizzati per partita	2.539	.193	.965	13.145	.000
	Punti subiti per partita	-2.412	.211	-.841	-11.458	.000

1. Variabile dipendente: partite vinte

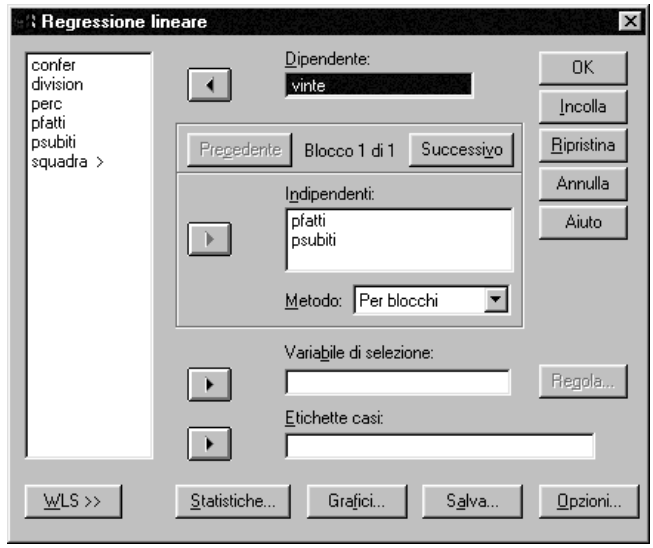


Per ottenere un'analisi della regressione lineare

▶ Dai menu, scegliere:

Analizza
  Regressione
    Lineare...

Figura 27-2
Finestra di dialogo Regressione lineare



- ▶ Nella finestra di dialogo Regressione lineare, selezionare una variabile dipendente numerica.
- ▶ Selezionare una o più variabili indipendenti numeriche.

Se lo si desidera, è possibile:

- Raggruppare le variabili indipendenti in blocchi e specificare metodi di inserimento differenti per sottogruppi di variabili diversi.
- Scegliere una variabile di selezione per limitare l'analisi a un sottoinsieme di casi con valori particolari per questa variabile.
- Selezionare una variabile di casi per l'identificazione di punti nei grafici.
- Selezionare una variabile di ponderazione per WLS per un'analisi dei minimi quadrati ponderati.

Minimi quadrati ponderati (WLS). Consente di ottenere un modello dei minimi quadrati ponderati. I dati vengono ponderati in base al reciproco della loro varianza. Le osservazioni con varianza elevata hanno un peso minore nell'analisi rispetto a quelle con varianza ridotta.

Metodi di selezione della variabile di regressione lineare

La selezione del metodo consente di specificare come vengono inserite nell'analisi le variabili indipendenti. Utilizzando diversi metodi, è possibile creare molteplici modelli di regressione dallo stesso insieme di variabili.

- **Per blocchi (Regressione).** Metodo di selezione in cui tutte le variabili di un blocco sono inserite in un unico passo.
- **Per passi.** Ad ogni passo, la variabile indipendente non presente nell'equazione che ha la più bassa probabilità di F viene inserita, se quella probabilità è sufficientemente piccola. Le variabili già presenti nell'equazione di regressione vengono rimosse se la loro probabilità di F diviene sufficientemente elevata. Il metodo termina quando nessuna variabile rispetta il criterio di inserimento o quello di rimozione.
- **Rimozione (Selezione di variabili).** Una procedura per la selezione di variabili in cui tutte le variabili di un blocco sono rimosse in un solo passo.

- **Eliminazione all'indietro.** Una procedura di selezione di variabili nella quale tutte le variabili vengono inserite nell'equazione e poi rimosse sequenzialmente. La variabile con la più bassa correlazione parziale rispetto alla variabile dipendente viene considerata la prima da rimuovere. Essa viene rimossa se soddisfa il criterio di eliminazione. Dopo che la prima variabile è stata rimossa, la variabile con la più bassa correlazione parziale tra quelle rimaste nell'equazione, viene considerata come la prossima da eliminare. La procedura termina quando non ci sono più variabili nell'equazione che soddisfano il criterio di rimozione.
- **Selezione in avanti.** Una procedura di selezione delle variabili nella quale le variabili vengono inserite in modo sequenziale all'interno del modello. La prima variabile da inserire nell'equazione è quella con la più elevata correlazione positiva o negativa con la variabile dipendente. Questa variabile viene inserita nell'equazione solo se soddisfa il criterio di inserimento. Se è stata inserita la prima variabile, viene considerata come successiva la variabile indipendente non presente nell'equazione che ha la più elevata correlazione parziale. La procedura termina quando non ci sono più variabili che soddisfano il criterio di inserimento.

I valori di significatività dell'output si basano sull'adattamento di un singolo modello. Pertanto, i valori di significatività in genere non sono validi quando viene utilizzato un metodo stepwise (Per passi, Avanti o Indietro).

Tutte le variabili devono soddisfare il criterio di tolleranza per essere inserite nell'equazione, indipendentemente dal metodo di inserimento specificato. Il livello di tolleranza predefinito è 0,0001. Una variabile non viene inserita se può far sì che la tolleranza di un'altra variabile già nel modello non rientri nel criterio di tolleranza già stabilito.

Tutte le variabili indipendenti selezionate vengono aggiunte a un solo modello di regressione. È tuttavia possibile specificare diversi metodi di inserimento per diversi sottoinsiemi di variabili. Ad esempio, è possibile inserire un blocco di variabili nel modello di regressione utilizzando la selezione per passi e un secondo blocco utilizzando la selezione in avanti. Per aggiungere un secondo blocco di variabili a un modello di regressione, fare clic su Avanti.

Regressione lineare: Imposta regola

Figura 27-3
Finestra di dialogo Regressione lineare: Imposta regola

Nell'analisi verranno inseriti i casi definiti dalla regola di selezione impostata. Ad esempio, se si seleziona la variabile Uguale a e si immette 5 per il valore, nell'analisi verranno inclusi solo i casi in cui la variabile selezionata ha un valore uguale a 5. È inoltre possibile specificare un valore stringa.

Regressione lineare: grafici

Figura 27-4
Finestra di dialogo Regressione lineare: grafici

I grafici possono facilitare la validazione delle ipotesi di normalità, linearità e uguaglianza delle varianze. I grafici sono inoltre utili per la rilevazione di valori anomali, osservazioni insolite e casi di influenza. Dopo averli salvati come nuove variabili, sarà possibile utilizzare valori attesi, residui e altre diagnostiche disponibili

nell'Editor dei dati per la creazione di grafici con le variabili indipendenti. Sono disponibili i seguenti grafici:

Grafici a dispersione. Consentono di rappresentare due elementi qualsiasi tra i seguenti: la variabile dipendente, i valori attesi standardizzati, i residui standardizzati, i residui cancellati, i valori attesi corretti, i residui studentizzati o i residui cancellati studentizzati. Rappresentare nel grafico i residui standardizzati e i valori attesi standardizzati per verificare la linearità e l'uguaglianza delle varianze.

Elenco di variabili sorgente. Elenca la variabile dipendente (DEPENDNT) e le seguenti variabili (valori stimati e residui): valori stimati standardizzati (*ZPRED), residui standardizzati (*ZRESID), residui cancellati (*DRESID), valori stimati corretti (*ADJPRED), residui studentizzati (*SRESID), residui cancellati studentizzati (*SDRESID).

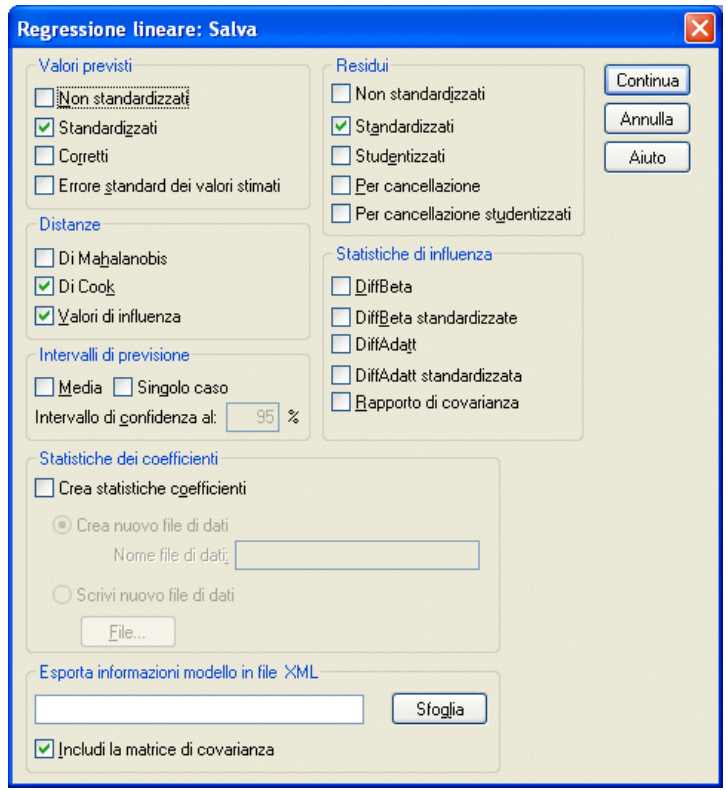
Produci tutti i grafici parziali. Consente di visualizzare i grafici a dispersione dei residui di ogni variabile indipendente e i residui della variabile dipendente quando entrambe le variabili sono regresse separatamente dal resto delle variabili indipendenti. Per la creazione di un grafico parziale è necessario che nell'equazione siano rappresentate almeno due variabili indipendenti.

Grafici dei residui standardizzati. Consentono di ottenere istogrammi dei residui standardizzati e grafici di probabilità normale confrontando la distribuzione dei residui standardizzati e la distribuzione normale.

Se vengono richiesti grafici, verranno visualizzate statistiche riassuntive per i valori attesi standardizzati e per i residui standardizzati (*ZPRED e *ZRESID).

Regressione lineare: Per salvare nuove variabili

Figura 27-5
Finestra di dialogo Regressione lineare: Salva



È possibile salvare valori attesi, residui e altre statistiche utili per la diagnostica. Ogni selezione aggiunge una o più nuove variabili al file dati attivo.

Valori attesi. I valori previsti dal modello di regressione per ogni caso.

- **Non standardizzati.** I valori risultanti dal modello per la variabile dipendente e per ciascun caso.
- **Standardizzati.** Una trasformazione di ciascun valore atteso nella sua forma standardizzata. Cioè, il valore atteso medio è sottratto dal valore atteso, e la differenza è divisa per la deviazione standard dei valori attesi. I valori attesi standardizzati hanno media 0 e deviazione standard 1.

- **Corretti.** Il valore atteso per un caso quando quel caso è escluso dal calcolo dei coefficienti di regressione.
- **E.S. dei valori stimati.** L'errore standard della media per i valori stimati. È una stima della deviazione standard del valor medio della variabile dipendente per i casi che assumono valori uguali nelle variabili indipendenti.

Distanze. Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere un notevole peso sul modello di regressione.

- **Mahalanobis.** Una misura della distanza di un caso dalla media di tutti i casi per le variabili indipendenti. Un valore alto indica che un caso include valori estremi per una o più variabili indipendenti.
- **Di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione cambierebbe sostanzialmente il loro valore.
- **Valori d'influenza.** Una misura dell'influenza di un dato sull'adattamento della regressione. L'influenza centrata varia da 0 (nessuna influenza sull'adattamento) a (N-1)/N.

Intervalli di previsione. I limiti superiore ed inferiore per gli intervalli di previsione singoli e medi.

- **Media (Intervallo di confidenza).** Limiti inferiore e superiore (due variabili) per l'intervallo di stima della risposta media.
- **Singolo.** Limiti inferiore e superiore (due variabili) per l'intervallo di stima della variabile dipendente per un singolo caso.
- **Intervallo di confidenza.** Inserire un valore fra 1 e 99,99 per specificare l'intervallo di confidenza per i due intervalli di stima. Occorre aver selezionato Media o Individuale per disporre di questa opzione. I valori tipici dell'intervallo di confidenza sono 90, 95 e 99.

Residui. Il valore effettivo della variabile dipendente meno il valore atteso dall'equazione di regressione.

- **Non standardizzati.** La differenza tra un valore osservato e il valore stimato dal modello.

- **Standardizzati.** Il residuo standardizzato, conosciuto anche come residuo di Pearson, è il residuo semplice diviso per l'errore standard della frequenza osservata. Benché i residui standardizzati siano asintoticamente normali, si comportano in modo meno variabile (in termini di varianza) delle variabili casuali normali. La loro distribuzione ha media 0 e deviazione standard 1
- **Studentizzati.** Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, in funzione della distanza tra ogni valore delle variabili indipendenti per un caso e la media di tutte le variabili indipendenti per lo stesso caso. Il residuo studentizzato riflette in maniera più precisa rispetto al residuo standardizzato le variazioni dell'errore da caso a caso.
- **Con rimozione.** Il residuo per un caso se quel caso venisse escluso dal calcolo dei coefficienti di regressione. È la differenza fra il valore della variabile dipendente e l valore stimato corretto.
- **Residui cancellati studentizzati.** Il residuo cancellato per un caso diviso per il suo errore standard. La differenza tra un residuo cancellato studentizzato ed il suo corrispondente residuo studentizzato indica quanta differenza produca l'eliminazione di un caso sulla stima del medesimo.

Statistiche di influenza. La modifica nei coefficienti di regressione (DiffBeta) e nei valori attesi (DiffAdatt) che risultano dall'esclusione di un particolare caso. I valori DiffBeta e DiffAdatt standardizzati sono anche disponibili con il rapporto di covarianza.

- **Differenza in beta.** Variazione del coefficiente di regressione quando un caso particolare viene eliminato dall'analisi. Viene calcolato un valore per ogni termine del modello, incluso il termine costante.
- **Differenza standardizzata in beta.** La variazione di un coefficiente di regressione quando un caso viene rimosso dall'analisi. Possono essere esaminati i casi con valore assoluto superiore a 2 diviso per la radice quadrata di N, dove N è il numero dei casi. Viene calcolato un valore per ogni termine del modello, incluso il termine costante.
- **Differenza nel valore adattato.** Variazione del valore adattato quando un caso viene escluso dall'analisi.

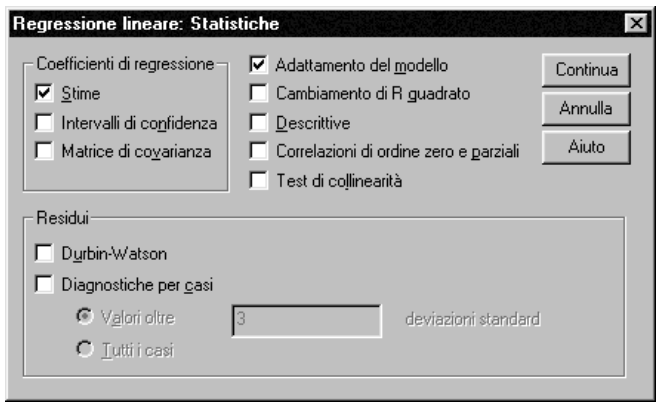
- **Differenza standardizzata nel valore adattato.** La variazione del valore stimato quando un caso viene rimosso dall'analisi. Si possono esaminare valori maggiori in valore assoluto a 2 diviso la radice quadrata di p/N, dove p è il numero di variabili indipendenti nell'equazione e N è il numero di casi.
- **Rapporto di covarianza.** Rapporto fra il determinante della matrice di covarianza con un caso escluso dal calcolo dei coefficienti di regressione ed il determinante con tutti i casi inclusi. Se il rapporto è vicino a 1, il caso non altera significativamente la matrice di covarianza.

Statistiche dei coefficienti. Salva i coefficienti di regressione in un file di dati. I file di dati possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali alla fine della sessione. I nomi dei file di dati devono essere conformi alle regole dei nomi delle variabili di SPSS. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.

Esporta informazioni modello in file XML. Le stime dei parametri e, se si desidera, le relative covarianze vengono esportati nel file specificato in formato XML (PMML). SmartScore e la versione del server di SPSS (un prodotto separato) possono utilizzare questo file di modello per applicare le informazioni del modello ad altri file dati a scopo di assegnazione dei punteggi.

Regressione lineare: Statistiche

Figura 27-6
Finestra di dialogo Regressione lineare: Statistiche



Sono disponibili le seguenti statistiche:

Coefficienti di regressione. Stime consente di visualizzare il coefficiente di regressione *B*, l'errore standard di *B*, il coefficiente beta standardizzato, il valore *t* per *B*e il livello di significatività a due code di *t*. Intervalli di confidenza consente di visualizzare gli intervalli di confidenza al 95% per ogni coefficiente di regressione o una matrice di covarianza. Matrice di covarianza consente di visualizzare una matrice di varianza-covarianza dei coefficienti di regressione con le covarianze esterne alla diagonale e le varianze sulla diagonale. Viene inoltre visualizzata una matrice di correlazione.

Adattamento del modello. Vengono elencate le variabili inserite ed eliminate dal modello e vengono visualizzate le seguenti statistiche di bontà dell'adattamento: *R* multipli, $R^{quadrato}$ e $R^{quadrato}$ corretto, errore standard della stima e una tabella di analisi della varianza.

Cambiamento di R quadrato. La variazione nella statistica $R^{quadrato}$ che viene prodotta aggiungendo o eliminando una variabile indipendente. Se la variazione di $R^{quadrato}$ associata ad una variabile è elevata, ciò significa che la variabile è un valido stimolatore della variabile dipendente.

Descrittive. Fornisce il numero di casi validi, la media e la deviazione standard per ogni variabile nell'analisi. Vengono inoltre visualizzati una matrice di correlazione con un livello di significatività a una coda e il numero di casi per ogni correlazione.

Correlazione parziale. Ciò che rimane della correlazione fra due variabili, dopo aver rimosso gli effetti della loro reciproca correlazione con altre variabili. Per esempio, la correlazione fra la variabile dipendente e una variabile indipendente, dopo aver rimosso gli effetti della correlazione fra la variabile indipendente e le altre variabili indipendenti incluse nel modello.

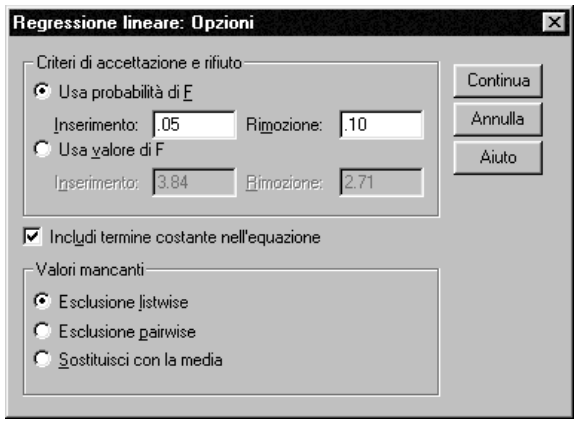
Parziale (Regressione: output). La correlazione fra la variabile dipendente e una variabile indipendente dopo aver rimosso dall'indipendente gli effetti dovuti alla correlazione con le altre variabili indipendenti nel modello.

Diagnostiche di collinearità. La collinearità (o multicollinearità) è la situazione in cui una delle variabili indipendenti è una funzione lineare di altre variabili indipendenti. Consente di visualizzare gli autovalori della matrice dei prodotti incrociati scalata e non centrata, l'indice di collinearità e le proporzioni della decomposizione della varianza con i fattori di inflazione della varianza (VIF) e le tolleranze per le singole variabili.

Residui. Consente di visualizzare il test di Durbin-Watson per la correlazione seriale dei residui e la diagnostica per i casi che corrispondono al criterio di selezione (valori anomali superiori a *n* deviazioni standard).

Regressione lineare: Opzioni

Figura 27-7
Finestra di dialogo Regressione lineare: Opzioni



Sono disponibili le seguenti opzioni:

Criteri di accettazione e rifiuto. Queste opzioni vengono utilizzate quando si specifica il metodo di selezione delle variabili in avanti, indietro o per passi. È possibile inserire o eliminare le variabili dal modello in base alla significatività (probabilità) del valore *F* o allo stesso valore *F*.

- **Usa probabilità di F.** Le variabili vengono inserite nel modello se il livello di significatività del loro valore di F è minore di quello specificato come criterio di inserimento. Le variabili vengono altresì rimosse dal modello se il livello di significatività del loro valore di F è maggiore di quello specificato come criterio di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi, e Inserimento deve essere minore di Rimozione. Alzando il valore di inserimento e/o abbassando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.
- **Usa valore di F.** Le variabili vengono inserite nel modello se il loro valore di F è maggiore di quello specificato come criterio di inserimento. Le variabili vengono altresì rimosse dal modello se il loro valore di F è minore di quello specificato

come criterio di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi, e Inserimento deve essere maggiore di Rimozione. Abbassando il valore di inserimento e/o alzando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.

Includi termine costante nell'equazione. Per impostazione predefinita, il modello di regressione include un termine costante. Se l'opzione è deselezionata, viene forzato il passaggio della curva di regressione per l'origine, il che avviene raramente. Alcuni risultati di una curva di regressione che passa per l'origine non sono confrontabili con i risultati della regressione che include una costante. Ad esempio *R*quadrato non può essere interpretato nel modo usuale.

Valori mancanti. È possibile scegliere tra le opzioni seguenti:

- **Esclusione listwise.** Sono inclusi nell'analisi solo i casi con valori validi per tutte le variabili.
- **Esclusione pairwise.** Per calcolare il coefficiente di correlazione su cui si basa l'analisi della regressione vengono utilizzati i casi con dati completi per la coppia di variabili correlate. I gradi di libertà sono basati su *N* pairwise minimo.
- **Sostituisci con la media.** Per i calcoli vengono utilizzati tutti i casi e la media della variabile viene sostituita alle osservazioni mancanti.

Opzioni aggiuntive del comando REGRESSION

Il linguaggio a comandi consente inoltre di:

- Scrivere una matrice di correlazione o leggere una matrice anziché i dati grezzi per ottenere l'analisi di regressione (con il sottocomando `MATRIX`).
- Specificare i livelli di tolleranza (tramite il sottocomando `CRITERIA`).
- Ottenere più modelli per variabili dipendenti uguali o diverse (con i sottocomandi `METHOD` e `DEPENDENT`).
- Ottenere statistiche aggiuntive (con i sottocomandi `DESCRIPTIVES` e `STATISTICS`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
28

Stima di curve

La procedura Stima di curve produce le statistiche di regressione per la stima di curve e i grafici correlati per 11 diversi modelli di regressione per la stima di curve. Per ciascuna variabile dipendente verrà creato un modello distinto. È inoltre possibile salvare come nuove variabili i valori attesi, i residui e gli intervalli di stima.

Esempio. Un provider di servizi Internet deve tener traccia della percentuale di traffico e-mail infettato da virus sulla propria rete nell'arco di un periodo di tempo specifico. Il grafico di dispersione indica che la relazione non è lineare. È necessario adattare ai dati un modello quadratico o cubico e controllare la validità delle assunzioni e la bontà di adattamento del modello.

Statistiche. Per ciascun modello: coefficienti di regressione, *R* multiplo, R^{quadrato}, R^{quadrato} corretto, errore standard della stima, tabella di analisi della varianza, valori attesi, residui ed intervalli di stima. Modelli: lineare, logaritmico, inverso, quadratico, cubico, di potenza, composto, curva S, logistico, di crescita ed esponenziale.

Dati. Le variabili dipendenti ed indipendenti devono essere quantitative. Se si seleziona Tempo come variabile indipenente dal file di dati attivo anziché selezionare una variabile, la procedura Stima di curve genera una variabile tempo se il periodo di tempo tra i casi è uniforme. Se si seleziona Tempo, la variabile dipendente deve essere una serie storica. L'analisi di serie storiche richiede nel file di dati una struttura in cui ciascun caso (riga) rappresenta una serie di osservazioni eseguite a orari diversi e il periodo di tempo fra i casi è uniforme.

Assunzioni. Valutare i dati graficamente per determinare il tipo di relazione esistente tra variabili dipendenti e indipendenti (lineare, esponenziale, ecc.). I residui di un buon modello devono essere normali e distribuiti casualmente. Se si utilizza un modello lineare, devono essere soddisfatte le seguenti ipotesi. Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. La relazione tra la variabile

dipendente e la variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti.

Figura 28-1
Tabella riassuntiva della stima delle curve

R	R-quadrato	R-quadrato corretto	Errore standard della stima
,885	,783	,766	,083

La variabile indipendente è Ore trascorse dalla rilevazione.

Figura 28-2
Stima delle curve ANOVA

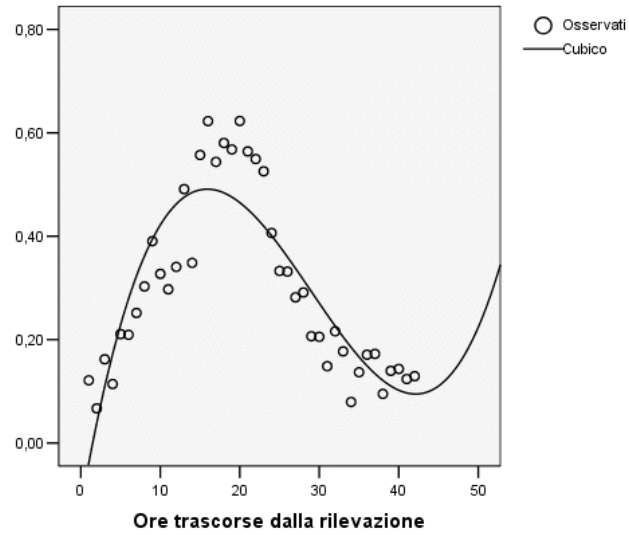
	Somma dei quadrati	df	Media dei quadrati	F	Sig.
Regressione	,949	3	,316	45,736	,000
Residuo	,263	38	,007		
Totale	1,212	41			

La variabile indipendente è Ore trascorse dalla rilevazione.

Figura 28-3
Coefficienti della stima delle curve

	Coefficienti non standardizzati		Coefficienti standardizzati		
	B	Errore standard	Beta	t	Sig.
Ore trascorse dalla rilevazione	,088	,011	6,305	7,888	,000
Ore trascorse dalla rilevazione ** 2	-,004	,001	-12,111	-6,363	,000
Ore trascorse dalla rilevazione ** 3	4,40E-005	,000	5,638	4,780	,000
(Costante)	-,123	,056		-2,187	,035

Figura 28-4
Grafico della stima delle curve

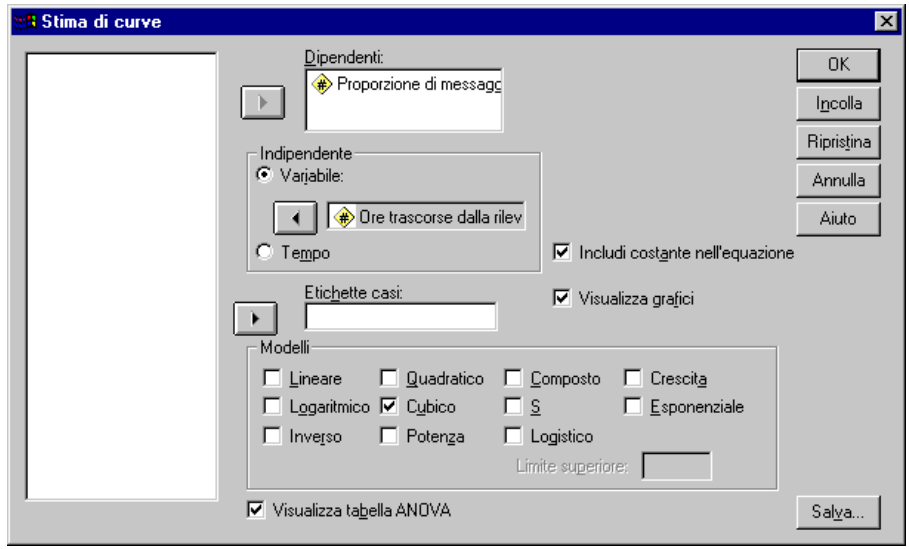


Per ottenere una stima di curve

▶ Dai menu, scegliere:

Analizza
  Regressione
    Stima di curve...

Figura 28-5
Finestra di dialogo Stima di curve



- Selezionare una o più variabili dipendenti. Per ciascuna variabile dipendente verrà creato un modello distinto.
- Selezionare una variabile indipendente (una variabile nel file dati attivo oppure Tempo).
- Oppure:
  - Selezionare una variabile per etichettare casi nei grafici a dispersione. Per ciascun punto del grafico a dispersione, è possibile utilizzare lo strumento di selezione dei punti per visualizzare il valore della variabile Etichetta di caso.
  - Fare clic su Salva per salvare i valori attesi, i residui e gli intervalli di stima come nuove variabili.

Sono inoltre disponibili le seguenti opzioni:

- **Includi termine costante nell'equazione.** Consente di valutare un termine costante nell'equazione di regressione. La costante viene inclusa per default.

- **Visualizza grafici.** Consente di tracciare i valori della variabile dipendente e ciascun modello selezionato in base alla variabile indipendente. Viene prodotto un grafico per ogni variabile dipendente.
- **Visualizza tabella ANOVA.** Consente di visualizzare una tabella di analisi della varianza per ciascun modello selezionato.

Stima di curve: Modelli

È possibile scegliere uno o più modelli di regressione per la stima di curve. Per determinare il modello da utilizzare, tracciare i dati in un grafico. Se le variabili appaiono legate da una relazione lineare, utilizzare un modello di regressione lineare semplice. Se la relazione tra le variabili non è lineare, provare a trasformare i dati. Se la trasformazione non risulta utile, può essere necessario utilizzare un modello più complesso. Visualizzare i dati in un grafico a dispersione; se il grafico è simile a una funzione matematica nota, adattare i dati a quel tipo di modello. Se, ad esempio, i dati sono simili a una funzione esponenziale, utilizzare il modello esponenziale.

Lineare. Modello la cui equazione è Y=b0+(b1*t). I valori della serie sono rappresentati come una funzione lineare del tempo.

Logaritmico. Modello la cui equazione è Y=b0+(b1*ln(t)).

Inverso. Modello la cui equazione è Y=b0+(b1/t).

Quadratico. Modello la cui equazione è: Y=b0+(b1*t)+(b2*t**2). Il modello quadratico può essere usato per modellare una serie che "decolla" o una serie che si smorza rapidamente.

Cubico. Modello definito dall'equazione: Y=b0+(b1*t) + (b2*t**2) + (b3*t**3)

Potenza. Modello la cui equazione è: Y=b0*(t**b1) oppure ln(Y)=ln(b0)+(b1*ln(t)).

Composto. Modello la cui equazione è Y=b0*(b1**t) oppure ln(Y)=ln(b0)+(ln(b1)*t)

Curva S. Modello la cui equazione è Y=e**(b0+(b1/t)) oppure ln(Y)=b0+(b1/t).

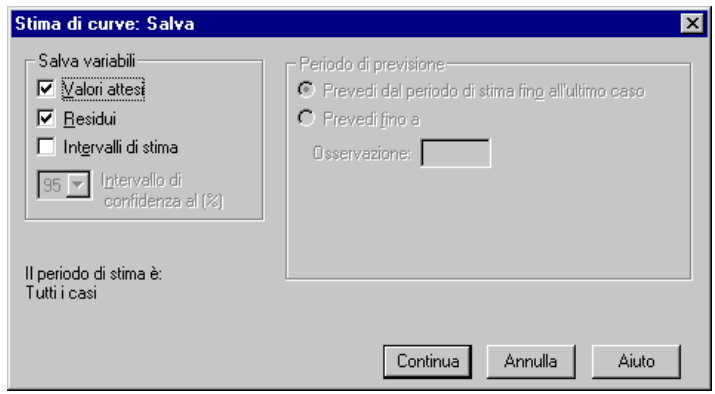
Logistico. Modello la cui equazione è Y=1/(1/u+(b0*(b1**t))) oppure ln(1/Y-1/u)=ln(b0+(ln(b1)*t)) dove u è il limite superiore. Specificare il limite superiore, che deve essere un intero positivo maggiore del valore più alto assunto dalla variabile dipendente. Il valore deve essere un numero positivo maggiore del valore più alto della variabile dipendente.

Di crescita. Modello la cui equazione è Y=e**(b0+(b1*t)) oppure ln(Y)=b0+(b1*t).

Esponenziale. Modello la cui equazione è Y=b0*(e**(b1*t)) oppure ln(Y)=ln(b0)+(b1*t).

Stima di curve: Salva

Figura 28-6
Finestra di dialogo Stima di curve: Salva



Salva Variabili. Per ciascun modello selezionato è possibile salvare i valori attesi, i residui (valori osservati della variabile dipendente meno il valore atteso del modello) e gli intervalli di stima (limite superiore e inferiore). I nomi delle nuove variabili e le etichette descrittive vengono visualizzati in una tabella nell'output.

Periodo di previsione. Se come variabile indipendente si seleziona Tempo anzichè una variabile nel file dati attivo, è possibile specificare un periodo di previsione al termine delle serie storiche. È possibile scegliere una delle seguenti opzioni:

- **Prevedi dal periodo di stima fino all'ultimo caso.** Consente di prevedere i valori per tutti i casi del file in base ai casi inclusi nel periodo di stima. Il periodo di stima, visualizzato in fondo alla finestra di dialogo, viene definito nella sottofinestra di dialogo dell'opzione Seleziona casi del menu Dati. Se non è stato definito alcun periodo di stima, i valori verranno previsti in base a tutti i casi.
- **Prevedi fino a.** Consente di prevedere i valori fino alla data, all'ora o al numero di osservazione specificato in base ai casi inclusi nel periodo di stima. Questa funzione può essere usata per prevedere valori futuri nelle serie storiche. Le variabili data definite specificano quali caselle di testo è possibile usare per specificare la fine di un periodo di stima. Se non vengono definite variabili di data, è possibile specificare l'ultimo numero di osservazione (caso).

Per creare variabili di data, utilizzare l'opzione Definisci date, disponibile nel menu Dati.

Capitolo
29

Analisi discriminante

L'Analisi discrimante è utile per creare modelli predittivi di gruppi di appartenenza basati sulle caratteristiche osservate di ciascun caso. La procedura genera una funzione discriminante oppure, per più di due gruppi, un insieme di funzioni discriminanti, in base alle combinazioni lineari delle variabili stimatore che forniscono la migliore discriminazione tra i gruppi. Le funzioni vengono generate da un campione di casi di cui è noto il gruppo di appartenenza; è quindi possibile applicare le funzioni ai nuovi casi con misurazioni per le variabili stimatore, ma di cui non è noto il gruppo di appartenenza.

Nota: La variabile di raggruppamento può includere più di due valori. I codici per la variabile di raggruppamento devono tuttavia essere interi, ed è necessario specificare i valori massimo e minimo corrispondenti. I casi con valore non compreso tra i due estremi specificati vengono esclusi dall'analisi.

Esempio. In media, il consumo calorico giornaliero degli abitanti delle zone temperate è maggiore di quello di chi vive ai tropici. Nelle zone temperate, inoltre, si riscontra una maggiore percentuale di persone che vivono in ambiente urbano. Un ricercatore desidera combinare queste informazioni in una funzione per determinare le modalità di discriminazione tra i due gruppi di paesi. Il ricercatore ritiene opportuno prendere in considerazione anche le dimensioni della popolazione e informazioni di carattere economico. L'analisi discriminante consente di valutare i coefficienti della funzione discriminante lineare, analoga alla parte destra di un'equazione di regressione lineare multipla. In altri termini, utilizzando i coefficienti *a*, *b*, *c* e *d* si ottiene la funzione:

D = a * clima + b * urbano + c * popolazione + d * prodotto interno lordo pro capite

Se queste variabili sono utili per la discriminazione tra le due zone climatiche, i valori di *D* per i paesi temperati saranno diversi da quelli relativi ai paesi tropicali. Se è necessario usare un metodo di selezione delle variabili per passi, nella funzione non si dovranno includere tutte e quattro le variabili.

Statistiche. Per ogni variabile: media, deviazione standard, ANOVA univariata. Per ogni analisi: *M* di Box, matrice di correlazione entro gruppi, matrice di covarianza entro gruppi, matrice di covarianza di gruppi separati e matrice di covarianza totale. Per ciascuna funzione discriminante canonica: autovalori, percentuale di varianza, correlazione canonica, lambda di Wilks, chi-quadrato. Per ogni passo: probabilità a priori, coefficienti di funzione di Fisher, coefficienti di funzione standardizzati, lambda di Wilks per ogni funzione canonica.

Dati. La variabile di raggruppamento deve includere un numero limitato di categorie distinte, codificate come interi. Le variabili indipendenti nominali devono essere ricodificate in forma di variabili fittizie o di contrasto.

Assunzioni. I casi devono essere indipendenti. Le variabili stimatore devono avere una distribuzione normale multivariata e le matrici di varianza-covarianza entro gruppi devono essere uguali in tutti i gruppi. Si assume che le appartenenze ai gruppi si escludano reciprocamente (ovvero che nessun caso appartenga a più gruppi) e che ciascun caso appartenga a un gruppo. La procedura è più efficace se l'appartenenza ai gruppi è una variabile categoriale effettiva; se l'appartenenza ai gruppi si basa sui valori di una variabile continua (ad esempio, QI massimo e QI minimo), è opportuno utilizzare la regressione lineare per avvalersi delle informazioni più dettagliate disponibili nella variabile continua.

Figura 29-1
Output dell'analisi discriminante

Autovalori

Funzione	Autovalore	% di varianza	% cumulata	Correlazione canonica
1	1.002	100.0	100.0	.707

Lambda di Wilks

Test di funzioni	Lambda di Wilks	Chi-quadrato	df	Sig.
1	,499	31,934	4	,000

Matrice di struttura

	Funzione
	1
CALORIE	.986
LOGPIL	.790
URBANA	.488
LOGPOP	.082

Funzioni ai baricentri di gruppo

Tipologia climatica	Funzione
	1
tropicale	-.869
temperato	1.107

Per ottenere un'analisi discriminante

- Dai menu, scegliere:

  Analizza
  Classificazione
  Discriminante...

Figura 29-2
Finestra di dialogo Analisi discriminante

- Selezionare una variabile di raggruppamento con valori interi e fare clic su Definisci intervallo per specificare le categorie desiderate.

- Selezionare le variabili indipendenti o le variabili stimatore. (Se la variabile di raggruppamento non include valori interi, utilizzando il comando Ricodifica automatica del menu Trasforma è possibile crearne una che includa tali valori).

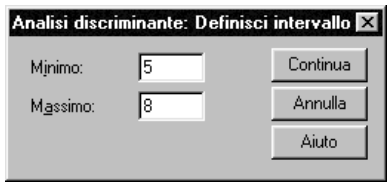
- Selezionare il metodo di inserimento delle variabili indipendenti.

- **Inserisci indipendenti insieme.** Inserisce contemporaneamente tutte le variabili indipendenti che soddisfano i criteri di tolleranza.
- **Usa metodo stepwise.** Usa un metodo per passi per controllare l'inserimento e la rimozione delle variabili.

▶ È inoltre possibile selezionare i casi utilizzando una variabile di selezione.

Analisi discriminante: Definisci intervallo

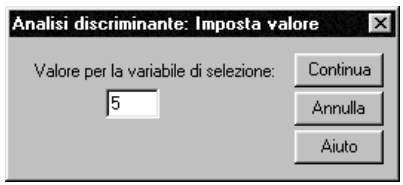
Figura 29-3
Finestra di dialogo Analisi discriminante: Definisci intervallo



Specificare il valore minimo e massimo della variabile di raggruppamento da utilizzare per l'analisi. I casi con valori che non rientrano in tale intervallo non vengono utilizzati nell'analisi discriminante, ma vengono classificati in uno dei gruppi esistenti in base ai risultati dell'analisi stessa. I valori massimo e minimo devono essere interi.

Analisi discriminante: Seleziona casi

Figura 29-4
Finestra di dialogo Analisi discriminante: Imposta valore



Per selezionare i casi da usare nell'analisi:

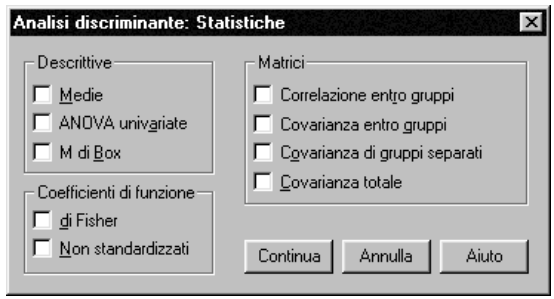
▶ Scegliere la variabile di selezione nella finestra di dialogo Analisi discriminante.

▶ Fare clic su Valore per immettere la variabile di selezione come numero intero.

Le funzioni discriminanti verranno derivate solo in base ai casi che prevedono tale valore per la variabile di selezione. Le statistiche e i risultati della classificazione vengono generati sia per i casi selezionati che per i casi non selezionati. Tale processo consente di classificare i nuovi casi sulla base dei dati preesistenti nonché ripartire i dati in sottoinsiemi di prova con i quali eseguire la convalida del modello generato.

Analisi discriminante: Statistiche

Figura 29-5
Finestra di dialogo Analisi discriminante: Statistiche



Descrittive. Le opzioni disponibili sono medie (incluse le deviazioni standard), ANOVA univariate e test *M* di Box.

- **Medie (Analisi discriminante).** Visualizza le medie e le deviazioni standard globali e di gruppo delle variabili indipendenti.
- **ANOVA univariate (Analisi discriminante).** Effettua l'analisi della varianza univariata per verificare l'ipotesi di uguaglianza delle medie di gruppo. Viene effettuata un'analisi per ciascuna variabile indipendente.
- **test M di Box.** Un test per l'uguaglianza di matrici di covarianza di gruppo. Per dimensioni di campione sufficientemente elevate, un valore P non significativo vuol dire che non ci sono sufficienti prove che le matrici differiscano. Il test è sensibile a scostamenti dalla normalità multivariata. Cioè, tende a non considerare uguali le matrici se l'ipotesi di normalità viene violata.

Coefficienti di funzione. Le opzioni disponibili sono i coefficienti di correlazione di Fisher e i coefficienti non standardizzati.

- **Di Fisher (Analisi discriminante).** Visualizza i coefficienti di Fisher della funzione discriminante, che possono essere usati direttamente per la classificazione. Viene riprodotto un insieme di coefficienti per ciascun gruppo. Ogni caso viene assegnato al gruppo in cui ottiene il più alto punteggio discriminante.
- **Non standardizzati (Analisi discriminante).** I coefficienti della funzione discriminante non vengono standardizzati.

Matrici. Le matrici di coefficienti disponibili per le variabili indipendenti sono: matrice di correlazione entro gruppi, matrice di covarianza entro gruppi, matrice di covarianza gruppi separati e matrice di covarianza totale.

- **Correlazione entro gruppi.** Visualizza la matrice di correlazione entro gruppi ottenuta mediando le matrici di covarianza di tutti i gruppi prima di calcolare le correlazioni.
- **Covarianza entro gruppi.** Visualizza una matrice combinata di covarianza entro i gruppi, mediando le singole matrici di covarianza di tutti i gruppi. Questa matrice può essere diversa dalla matrice di covarianza globale.
- **Covarianza di gruppi separati.** Visualizza le matrici di covarianza dei singoli gruppi.
- **Covarianza totale.** Visualizza la matrice di covarianza globale, calcolata su tutti i casi, ovvero ignorando la suddivisione in gruppi.

Analisi discriminante: Metodo Stepwise

Figura 29-6
Finestra di dialogo Analisi discriminante: Metodo Stepwise

Metodo. Selezionare la statistica da utilizzare per inserire o rimuovere nuove variabili. Le alternative disponibili sono: lambda di Wilks, varianza non spiegata, distanza di Mahalanobis, minimo rapporto *F* e *V* di Rao. Con il *V* di Rao è possibile specificare l'aumento minimo di *V* per la variabile da inserire.

- **Lambda di Wilks.** Un metodo di selezione delle variabili nell'analisi discriminante per passi che sceglie le variabili da inserire nel modello in base a quanto esse contribuiscono a minimizzare il Lambda di Wilks.
- **Varianza non spiegata.** Ad ogni passo viene inserita la variabile con maggior potere esplicativo della variazione fra gruppi, ovvero quella che minimizza la varianza non spiegata.
- **Distanza di Mahalanobis.** Una misura della distanza di un caso dalla media di tutti i casi per le variabili indipendenti. Un valore alto indica che un caso include valori estremi per una o più variabili indipendenti.
- **Minimo rapporto F.** Un metodo di selezione delle variabili nelle analisi per passi basato sulla massimizzazione di un rapporto F valutato tramite la distanza di Mahalanobis tra gruppi.
- **V di Rao.** Una misura delle differenze tra medie di gruppo. È anche detta traccia di Lawley-Hotelling. Ad ogni passo viene inserita la variabile che massimizza l'aumento della V di Rao. Specificare l'incremento minimo che una variabile deve apportare per essere inserita nel modello.

Criteri. Le alternative disponibili sono: Usa valore di F e Usa probabilità di F. Immettere valori per aggiungere o rimuovere variabili.

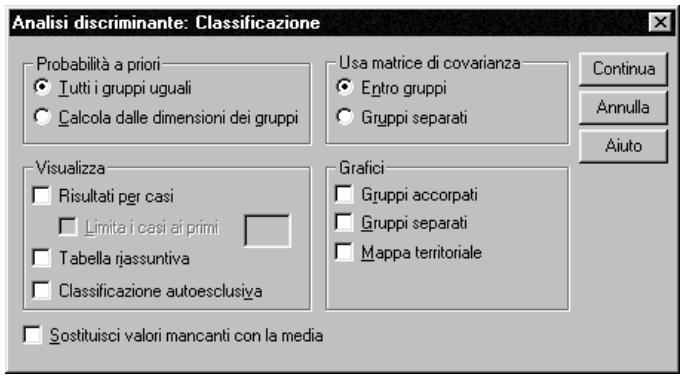
- **Usa valore di F.** Le variabili vengono inserite nel modello se il loro valore di F è maggiore di quello specificato come criterio di inserimento. Le variabili vengono altresì rimosse dal modello se il loro valore di F è minore di quello specificato come criterio di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi, e Inserimento deve essere maggiore di Rimozione. Abbassando il valore di inserimento e/o alzando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.
- **Usa probabilità di F.** Le variabili vengono inserite nel modello se il livello di significatività del loro valore di F è minore di quello specificato come criterio di inserimento. Le variabili vengono altresì rimosse dal modello se il livello di significatività del loro valore di F è maggiore di quello specificato come criterio di rimozione. I valori di inserimento e di rimozione devono essere entrambi positivi, e Inserimento deve essere minore di Rimozione. Alzando il valore

di inserimento e/o abbassando quello di rimozione si allentano i vincoli di inclusione delle variabili nel modello.

Visualizzazione. Riepilogo dei passi visualizza le statistiche per tutte le variabili dopo ogni passo. F per distanze a coppie visualizza una matrice di valori *F* pairwise per ciascuna coppia di gruppi.

Analisi discriminante: Classificazione

Figura 29-7
Finestra di dialogo Analisi discriminante: Classificazione



Probabilità a priori. Questi valori vengono utilizzati per la classificazione. È possibile specificare una probabilità a priori per tutti i gruppi (Tutti i gruppi uguali) oppure fare in modo che le probabilità di appartenenza ai gruppi vengano determinate in base alle dimensioni del gruppo osservato nel campione (Calcola dalle dimensioni dei gruppi).

Visualizzazione. Le opzioni di visualizzazione disponibili sono: risultati per casi, tabella riassuntiva e classificazione autoesclusiva.

- **Risultati per casi.** Produce per ciascun caso il vero gruppo di appartenenza, il gruppo previsto dal modello, la probabilità a posteriori e il punteggio discriminante.
- **Tabella riassuntiva.** Il numero di casi assegnati in modo corretto o non corretto a ciascuno dei gruppi dall'analisi discriminante. A volte detta "Matrice confusione".
- **Classificazione autoesclusiva.** Ogni caso viene classificato usando le funzioni costruite su tutti i casi meno se stesso. Nota anche come classificazione leave one out o metodo U.

Sostituisci valori mancanti con la media. Selezionare questa opzione per sostituire un valore mancante con la media di una variabile indipendente, solo durante la fase di classificazione.

Usa matrice di covarianza. È possibile classificare i casi utilizzando una matrice di covarianza entro gruppi o una matrice di covarianza gruppi separati.

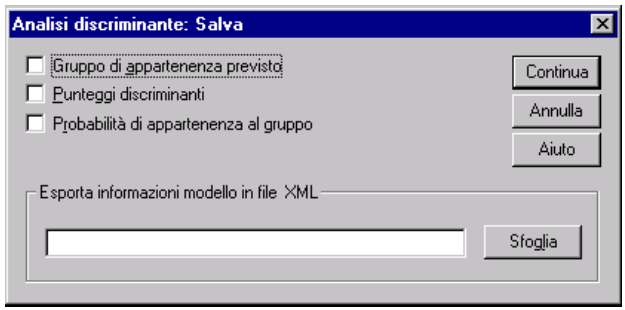
- **Entro gruppi.** Per classificare i casi viene utilizzata la matrice globale di covarianza entro i gruppi.
- **Gruppi separati.** Per classificare i casi vengono utilizzate le matrici di covarianza dei singoli gruppi gruppi. Dal momento che la classificazione è basata sulla funzione discriminante, e non sui valori originali, questa opzione non è sempre equivalente alla discriminazione quadratica.

Grafici. Le opzioni disponibili per i grafici sono: gruppi accorpati, gruppi separati e mappa territoriale.

- **Gruppi accorpati.** Produce un unico grafico a dispersione dei valori delle prime due funzioni disciminanti. Se c'è una sola funzione viene prodotto un istogramma.
- **Gruppi separati.** Produce un grafico a dispersione deii valori delle prime due funzioni disciminanti per ciascun gruppo. Se c'è una sola funzione vengono prodotti istogrammi.
- **Mappa territoriale.** Un grafico dei confini usati per classificare i casi in gruppi in base ai valori di una funzione. I numeri corrispondono ai gruppi nei quali i casi sono classificati. La media per ciascun gruppo è indicata da un asterisco all'interno dei suoi confini. La mappa non viene visualizzata se c'è una sola funzione discriminante.

Analisi discriminante: Salva

Figura 29-8
Finestra di dialogo Analisi discriminante: Salva



È possibile aggiungere nuove variabili al file di dati attivo. Le opzioni disponibili sono: gruppo di appartenenza previsto (una sola variabile), punteggi discriminanti (una variabile per ciascuna funzione discriminante nella soluzione) e probabilità di appartenenza al gruppo in base ai punteggi discriminanti (una variabile per ciascun gruppo).

È possibile esportare le informazioni sul modello nel file specificato in formato XML (PMML). SmartScore e la versione del server di SPSS (un prodotto separato) possono utilizzare questo file di modello per applicare le informazioni del modello ad altri file dati a scopo di assegnazione dei punteggi.

Funzioni aggiuntive del comando DISCRIMINANT

Il linguaggio a comandi SPSS permette anche di:

- Eseguire più analisi discriminanti (con un comando) e controllare l'ordine in cui le variabili vengono inserite (con il sottocomando `ANALYSIS`).
- Specificare le probabilità a priori per la classificazione (con il sottocomando `PRIORS`).
- Visualizzare le matrici dei motivi e delle strutture ruotate (con il sottocomando `ROTATE`).
- Limitare il numero di funzioni discriminanti estratte (con il sottocomando `FUNCTIONS`).
- Limitare la classificazione ai casi selezionati (o non selezionati) per l'analisi (con il sottocomando `SELECT`).

- Leggere e analizzare la matrice di correlazione (con il sottocomando `MATRIX`).
- Scrivere una matrice di correlazione da analizzare in seguito (con il sottocomando `MATRIX`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
30

Analisi fattoriale

L'analisi fattoriale si propone di identificare le variabili sottostanti, o **fattori**, che spiegano il modello di correlazioni all'interno di un insieme di variabili osservate. L'analisi fattoriale viene in genere utilizzata per la riduzione dei dati in quanto consente di identificare un numero ridotto di valori che spiegano la maggior parte dei valori di varianza osservati in numerose variabili manifeste. L'analisi fattoriale può inoltre essere utilizzata per generare ipotesi relative a meccanismi causali oppure per esaminare le variabili per le analisi successive (ad esempio per identificare la collinearità prima di eseguire un'analisi di regressione lineare).

La procedura di analisi fattoriale permette un elevato grado di flessibilità:

- Sono disponibili sette metodi di estrazione fattoriale.
- Sono disponibili cinque metodi di rotazione, tra cui oblimin diretto e promax per le rotazioni non ortogonali.
- Sono disponibili tre metodi per il calcolo dei punteggi, che possono essere salvati come variabili per le analisi successive.

Esempio. Quali sono gli atteggiamenti sottostanti che inducono le persone a rispondere a questionari politici in un determinato modo? Dall'esame delle correlazioni esistenti tra le voci del questionario risulta una significativa sovrapposizione tra diversi sottogruppi di voci. Ad esempio, le domande relative alle tasse tendono ad essere correlate fra loro, così come le domande relative alle questioni militari e così via. Grazie all'analisi fattoriale è possibile identificare il numero di fattori sottostanti e in molti casi determinare cosa rappresentano concettualmente tali fattori. È inoltre possibile calcolare i punteggi fattoriali per ciascun rispondente, un elemento che è possibile utilizzare in analisi successive. Ad esempio, è possibile creare un modello di regressione logistico che consenta di prevedere il comportamento di voto in base ai punteggi fattoriali.

Statistiche. Per ogni variabile: numero di casi validi, media e deviazione standard. Per ciascuna analisi fattoriale: matrice di correlazione delle variabili, inclusi i livelli di significatività, determinante, inversa; matrice di correlazione riprodotta, inclusa anti-immagine; soluzione iniziale (comunalità, autovalori e percentuale di varianza spiegata); misura di adeguatezza campionaria di Kaiser-Meyer-Olkin e test di sfericità di Bartlett; soluzione non ruotata, inclusi pesi fattoriali, comunalità e autovalori; soluzione ruotata, incluse la matrice ruotata dei modelli e la matrice di trasformazione. Per le rotazioni oblique: matrice ruotata dei modelli e delle strutture; matrice dei coefficienti di punteggio fattoriale e matrice di covarianza fattoriale. Grafici: grafico decrescente degli autovalori e grafico dei pesi fattoriali dei primi due o tre fattori.

Dati. Le variabili devono essere quantitative per **intervallo** oppure **intervallo** livello. I dati categoriali (ad esempio religione o paese d'origine) non sono idonei per l'analisi fattoriale. I dati per cui è possibile calcolare i coefficienti di correlazione di Pearson sono idonei per l'analisi fattoriale.

Assunzioni. I dati devono avere una distribuzione normale bivariata per ciascuna coppia di variabili e le osservazioni devono essere indipendenti. Il modello di analisi fattoriale specifica che le variabili vengono determinate da fattori comuni (i fattori stimati dal modello) e fattori univoci (che non risultano sovrapposti tra le variabili osservate); le stime calcolate si basano sull'ipotesi che tutti i fattori univoci siano correlati reciprocamente e con i fattori comuni.

Figura 30-1
Output dell'analisi fattoriale

Statistiche descrittive

	Media	Deviazione std.	Analisi fattoriale N
VITAFEM	72.83	8.27	72
MORTINF	35.132	32.222	72
ALFABETZ	82.47	18.63	72
TASSONAT	24.375	10.552	72
FERTILIT	3.205	1.593	72
URBANA	62.58	22.84	72
LOGPIL	3.5043	.6077	72
INCRPOP	1.697	1.156	72
NATSUMOR	3.5772	2.3131	72
TASSOMOR	8.04	3.17	72
LOGPOP	4.1526	.6865	72

Comunalità

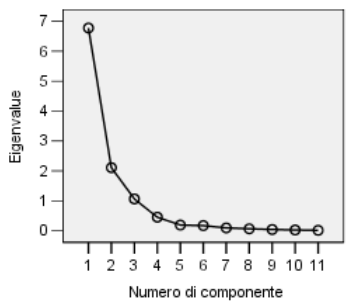
	Iniziale	Estrazione
VITAFEM	1.000	.953
MORTINF	1.000	.949
ALFABETZ	1.000	.825
TASSONAT	1.000	.943
FERTILIT	1.000	.875
URBANA	1.000	.604
LOGPIL	1.000	.738
INCRPOP	1.000	.945
NATSUMOR	1.000	.925
TASSOMOR	1.000	.689
LOGPOP	1.000	.292

Metodo di estrazione: Analisi componente principale.

Varianza totale spiegata

	Autovalori iniziali			Pesi dei fattori non ruotati			Pesi dei fattori ruotati		
Componente	Totale	% di varianza	% cumulata	Totale	% di varianza	% cumulata	Totale	% di varianza	% cumulata
1	6.242	56.750	56.750	6.242	56.750	56.750	6.108	55.525	55.525
2	2.495	22.685	79.435	2.495	22.685	79.435	2.630	23.910	79.435
3	.988	8.986	88.421						
4	.591	5.372	93.793						
5	.236	2.142	95.935						
6	.172	1.561	97.496						
7	.124	1.126	98.622						
8	.070	.633	99.254						
9	.045	.405	99.660						
10	.024	.222	99.882						
11	.013	.118	100.000						

Metodo di estrazione: Analisi componente principale.



Matrice di componente ruotata[1]

	Componente	
	1	2
TASSONAT	.969	
FERTILIT	.931	
ALFABETZ	-.880	.226
VITAFEM	-.856	.469
MORTINF	.853	-.469
INCRPOP	.847	.476
LOGPIL	-.794	.327
URBANA	-.561	.539
TASSOMOR		-.827
NATSUMOR	.614	.741
LOGPOP		-.520

Metodo estrazione: analisi componenti principali.
Metodo rotazione: Varimax con normalizzazione di kaiser.

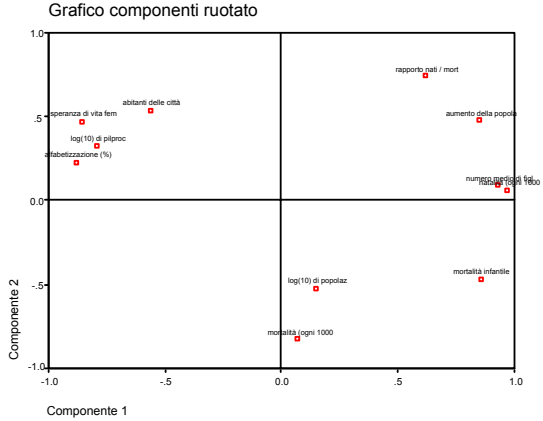
1. La rotazione ha raggiunto i criteri di convergenza in 3 iterazioni

Matrice di componente

		1	2
Componente	1	.982	-.190
	2	.190	.982

Metodo estrazione: analisi componenti principali.
Metodo rotazione: Varimax con normalizzazione di Kaiser.

Grafico componenti ruotato



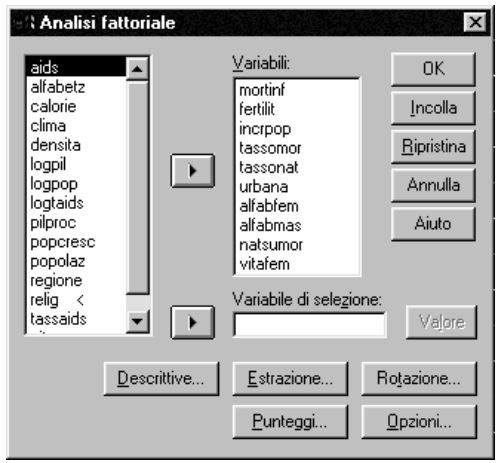
Per ottenere un'analisi fattoriale

- Dai menu, scegliere:

  Analizza
  Riduzione dati
  Fattore...

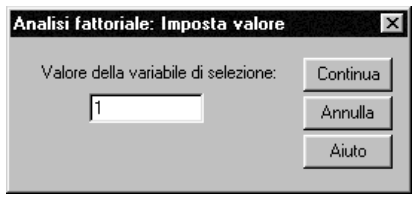
- Selezionare le variabili per l'analisi fattoriale.

Figura 30-2
Finestra di dialogo Analisi fattoriale



Analisi fattoriale: Seleziona casi

Figura 30-3
Finestra di dialogo Analisi fattoriale: Imposta valore



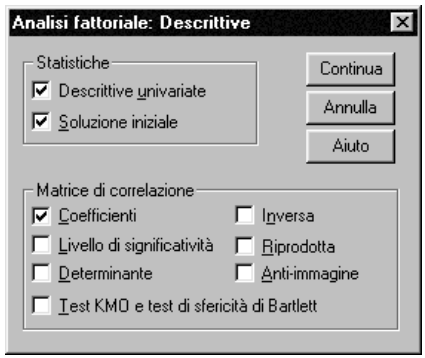
Per selezionare i casi da usare nell'analisi:

- Selezionare una variabile di selezione.
- Fare clic suValore per immettere la variabile di selezione come numero intero.

Nell'analisi fattoriale verranno utilizzati solo i casi con tale valore per la variabile di selezione.

Analisi fattoriale: Descrittive

Figura 30-4
Finestra di dialogo Analisi fattoriale: Descrittive



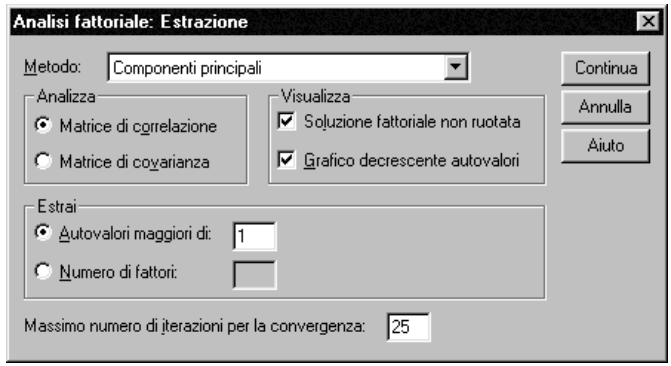
Statistiche. Le stastiche descrittive univariate includono la media, la deviazione standard e il numero di casi validi per ciascuna variabile. Nella soluzione iniziale vengono visualizzate le comunalità iniziali, gli autovalori e la percentuale della varianza spiegata.

Matrice di correlazione. Le opzioni disponibili sono: coefficienti, livelli di significatività, determinante, inversa, riprodotta, anti-immagine, test KMO e test di sfericità di Bartlett.

- **Test KMO e test di sfericità di Bartlett.** La misura di adeguatezza campionaria KMO (Keiser Meyer Olkin) verifica se le correlazioni parziali tra le variabili sono piccole. Il test di sfericità di Bartlett verifica se la matrice di correlazione è una matrice identità, cosa che indicherebbe l'inadeguatezza del modello fattoriale.
- **Riprodotta.** La matrice di correlazione stimata a partire dalla soluzione del fattore. Sono visualizzati anche i residui (differenze tra correlazioni stimate e osservate).
- **Anti-immagine.** La matrice di correlazione anti-immagine contiene i coefficienti di correlazione parziale cambiati di segno e la matrice di covarianza anti-immagine contiene le covarianze parziali cambiate di segno. In un buon modello fattoriale la maggior parte degli elementi fuori dalla diagonale dovrebbe avere valori bassi. La misura di adeguatezza campionaria di una variabile è visualizzata sulla diagonale della matrice di correlazione anti-immagine.

Analisi fattoriale: Estrazione

Figura 30-5
Finestra di dialogo Analisi fattoriale: Estrazione



Metodo. Consente di specificare il metodo di estrazione dei fattori. I metodi disponibili sono: componenti principali, minimi quadrati non ponderati, minimi quadrati generalizzati, massima verosimiglianza, fattorizzazione dell'asse principale, fattorizzazione alfa e fattorizzazione immagine.

- **Componenti principali (Analisi fattoriale).** Metodo usato per formare combinazioni lineari non correlate delle variabili osservate. La prima componente spiega la parte più alta di variabilità. Le componenti successive spiegano porzioni di variabilità decrescenti e sono tutte non correlate fra loro. L'analisi delle componenti principali viene usata per ottenere la soluzione fattoriale iniziale. Può essere usata quando una matrice di correlazione è singolare.
- **Minimi quadrati non pesati (Analisi fattoriale).** Un metodo per l'estrazione dei fattori che minimizza la somma dei quadrati delle differenze tra la matrice di correlazione osservata e quella riprodotta, ignorando le diagonali.
- **Minimi quadrati generalizzati (Analisi fattoriale).** Un metodo di estrazione dei fattori che minimizza la somma dei quadrati delle differenze tra la matrice di correlazione osservata e la matrice di correlazione riprodotta. Le correlazioni sono ponderate tramite l'inverso della loro unicità, così che alle variabili con elevata unicità viene dato meno peso rispetto a quelle con unicità inferiore.
- **Massima verosimiglianza (Analisi fattoriale).** Un metodo per l'estrazione dei fattori che produce le stime dei parametri che più verosimilmente hanno prodotto la matrice di correlazione osservata, se il campione è estratto da una distribuzione

normale multivariata. Le correlazioni sono pesate tramite l'inverso dell'unicità delle variabili.

- **Fattorizzazione dell'asse principale (Analisi fattoriale).** Un metodo di estrazione dei fattori dalla matrice di correlazione originale con coefficienti di correlazione multipla al quadrato posti sulla diagonale come stime iniziali delle comunalità. Questi pesi di fattore vengono usati per stimare nuove comunalità che sostituiscono le vecchie stime sulla diagonale. Le iterazioni continuano fino a che le variazioni nelle comunalità da un'iterazione alla successiva soddisfano il criterio di convergenza per l'estrazione.
- **Alfa (Analisi fattoriale).** Un metodo per l'estrazione dei fattori che considera le variabili nell'analisi come un campione estratto dall'universo delle variabili che descrivono il fenomeno. Massimizza l'Alpha di Cronbach dei fattori.
- **Fattorizzazione immagine (Analisi fattoriale).** Metodo di estrazione dei fattori sviluppato da Guttman e basato sulla teoria dell'immagine. La parte comune della variabile, detta immagine parziale, viene definita come la sua regressione lineare sulle variabili rimanenti, piuttosto che in funzione di fattori ipotetici.

Analizza. Consente di specificare una matrice di correlazione o una matrice di covarianza.

- **Matrice di correlazione.** Può risultare utile se le variabili dell'analisi vengono misurate su scale diverse.
- **Matrice di covarianza.** Può risultare utile se si intende applicare l'analisi fattoriale a più gruppi con varianze diverse per ogni variabile.

Estrai. È possibile mantenere tutti i fattori con autovalori superiori al valore specificato oppure mantenere solo il numero di fattori specificato.

Visualizzazione. Consente di richiedere la soluzione fattoriale non ruotata e un grafico decrescente degli autovalori.

- **Soluzione non ruotata (Analisi fattoriale).** Visualizza la matrice dei pesi fattoriali, le comunalità e gli autovalori della soluzione non ruotata.
- **Grafico decrescente degli autovalori.** Un grafico della varianza associata a ciascun fattore. Viene usato per determinare quanti fattori debbano essere mantenuti. Tipicamente, il grafico mostra una decisa diminuizione di pendenza nel punto in cui entrano in gioco i fattori meno rilevanti.

Massimo numero di iterazioni per la convergenza. Consente di specificare il numero massimo di passaggi che l'algoritmo può eseguire per valutare la soluzione.

Analisi fattoriale: Rotazione

Figura 30-6
Finestra di dialogo Analisi fattoriale: Rotazione

Metodo. Consente di selezionare il metodo di rotazione fattoriale. I metodi disponibili sono: varimax, equamax, quartimax, oblimin diretto e promax.

- **Varimax (Analisi fattoriale).** Un metodo di rotazione ortogonale che minimizza il numero di variabili che caricano fortemente ciascun fattore. Semplifica l'interpretazione dei fattori.
- **Metodo oblimin diretto.** Un metodo di rotazione obliqua (non ortogonale). Quando delta vale 0 (il valore di default), le soluzioni sono per la maggior parte oblique. Quando delta diventa negativo e aumenta in valore assoluto, i fattori cominciano ad essere meno obliqui. Inserire un numero minore o uguale a 0,8.
- **Quartimax (Analisi fattoriale).** Un metodo di rotazione che rende minimo il numero di fattori necessari per spiegare ogni variabile. Semplifica l'interpretazione delle variabili osservate.
- **Equamax (Analisi fattoriale).** Un metodo di rotazione che minimizza sia il numero di variabili che pesano in modo elevato su un fattore, sia il numero di fattori necessari per spiegare una variabile. È una combinazione dei metodi Varimax e Quartimax.
- **Promax (Analisi fattoriale).** Una rotazione obliqua, che ammette la correlazione fra fattori. Può essere utile per file di grandi dimensioni, perché è più veloce del metodo Oblimin.

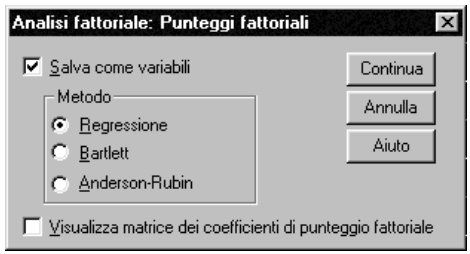
Visualizzazione. Consente di includere l'output nella soluzione ruotata, nonché i grafici dei pesi fattoriali per i primi due o tre fattori.

- **Soluzione ruotata (Analisi fattoriale).** Per ottenere la soluzione ruotata occorre aver selezionato un metodo di rotazione. Per le rotazioni ortogonali vengono visualizzate la matrice ruotata dei modelli e la matrice di trasformazione. Per le rotazioni oblique vengono visualizzate la matrice dei modelli, la matrice di struttura e la matrice di correlazione dei fattori.
- **Grafico dei pesi fattoriali.** Grafico tridimensionale dei pesi dei primi tre fattori. Per le soluzioni a due fattori viene prodotto un grafico bidimensionale. Assente se l'analisi estrae un solo fattore. Se è stata richiesta la rotazione, il grafico visualizza la soluzione ruotata.

Massimo numero di iterazioni per la convergenza. Consente di specificare il massimo numero di passaggi che l'algoritmo può eseguire per completare la rotazione.

Analisi fattoriale: Punteggi fattoriali

Figura 30-7
Finestra di dialogo Analisi fattoriale: Punteggi fattoriali



Salva come variabili. Consente di creare una nuova variabile per ciascun fattore nella soluzione finale. Selezionare uno dei seguenti metodi alternativi per calcolare i punteggi fattoriali: regressione, Bartlett o Anderson-Rubin.

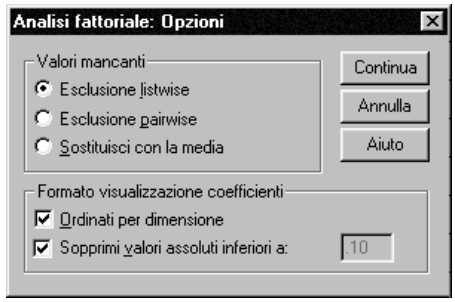
- **Metodo di regressione (Analisi fattoriale).** Un metodo per stimare i punteggi dei fattori. I punteggi prodotti hanno media 0 e varianza pari al quadrato della correlazione multipla fra i punteggi stimati ed i valori reali dei fattori. I punteggi possono essere correlati anche quando i fattori sono ortogonali.

- **Punteggi di Bartlett.** Un metodo di stima dei coefficienti di punteggio fattoriale. I punteggi fattoriali di Bartlett hanno media pari a 0. La somma dei quadrati dei singoli fattori sull'intervallo delle variabili è minimizzata.
- **Metodo di Anderson-Rubin.** Un metodo per calcolare i coefficienti dei punteggi fattoriali; rappresenta una modifica del metodo di Bartlett per assicurare l'ortogonalità dei fattori stimati. I punteggi forniti hanno una media pari a 0, deviazione standard pari ad 1 e risultano ortogonali (non correlati fra loro).

Visualizza matrice dei coefficienti di punteggio fattoriale. Mostra i coefficienti per cui vengono moltiplicate le variabili per ottenere i punteggi fattoriali. Vengono visualizzate anche le correlazioni tra i punteggi fattoriali.

Analisi fattoriale: Opzioni

Figura 30-8
Finestra di dialogo Analisi fattoriale: Opzioni



Valori mancanti. Consente di specificare le modalità di gestione dei valori mancanti. Le opzioni alternative disponibili sono: Esclusione **listwise**, Esclusione **pairwise**, o Sostituisci con la media.

Formato visualizzazione coefficienti. Consente di controllare alcuni aspetti delle matrici di output. È possibile ordinare i coefficienti per dimensioni ed eliminare i coefficienti con valori assoluti inferiori al valore specificato.

Opzioni aggiuntive del comando FACTOR

Il linguaggio a comandi SPSS permette anche di:

- Specificare i criteri di convergenza per l'iterazione durante l'estrazione e la rotazione.

- Specificare i singoli grafici dei fattori ruotati.
- Specificare quanti punteggi di fattori salvare.
- Specificare i valori diagonali per il metodo di calcolo dei fattori dell'asse principale.
- Scrivere le matrici di correlazione o le matrici dei fattori sul disco per poterle analizzare in seguito.
- Leggere e analizzare le matrici di correlazione o dei fattori.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
31
Scelta di una procedura di cluster analysis

Scelta di una procedura per il raggruppamento

È possibile eseguire cluster analysis utilizzando le procedure Analisi Cluster TwoStep, Cluster gerarchica o Cluster con metodo delle k-medie. Ogni procedura utilizza un algoritmo diverso per la creazione di cluster e include opzioni non disponibili nelle altre procedure.

Analisi Cluster TwoStep. La procedura Analisi Cluster TwoStep risulta appropriata per molte applicazioni. Rende disponibili le seguenti funzioni univoche:

- Selezione automatica del numero di cluster più appropriato, oltre a misure per la scelta tra modelli di cluster.
- Possibilità di creare contemporaneamente due modelli di cluster basati su variabili categoriali e continue.
- Possibilità di salvare il modello di cluster in un file XML esterno e quindi di leggere tale file e aggiornare il modello di cluster utilizzando i dati più recenti.

Inoltre, questa procedura consente di analizzare file di dati di grandi dimensioni.

Cluster gerarchica. La procedura Cluster gerarchica è limitata a file di dati di dimensioni minori (ad esempio per il raggruppamento di centinaia di oggetti) e rende disponibili le seguenti funzioni univoche:

- Possibilità di raggruppare casi o variabili in cluster.
- Possibilità di calcolare un intervallo di soluzioni possibili e di salvare l'appartenenza a un cluster per ognuna di queste soluzioni.
- Diversi metodi per la formazione di cluster, la trasformazione delle variabili e la misurazione della dissimilarità tra cluster.

La procedura Cluster gerarchica analizza variabili per intervallo (continue), binarie o di conteggio, purché le variabili siano tutte dello stesso tipo.

Cluster con metodo delle k-medie. La procedura Cluster con metodo delle k-medie è limitata a dati continui e richiede che il numero di cluster venga specificato anticipatamente, rendendo tuttavia disponibili le seguenti funzioni univoche:

- Possibilità di salvare le distanze dai centri di cluster per ogni oggetto.
- Possibilità di leggere centri iniziali di cluster da un file SPSS esterno e salvare centri finali di cluster in un file SPSS esterno.

Inoltre, questa procedura consente di analizzare file di dati di grandi dimensioni.

Capitolo
32

Analisi Cluster TwoStep

L'analisi Cluster TwoStep è uno strumento di esplorazione che consente di rilevare raggruppamenti naturali, o cluster, all'interno di insiemi di dati, che non sarebbero altrimenti evidenti. L'algoritmo utilizzato da questa procedura presenta diverse caratteristiche che lo differenziano dalle tecniche di raggruppamento tradizionali:

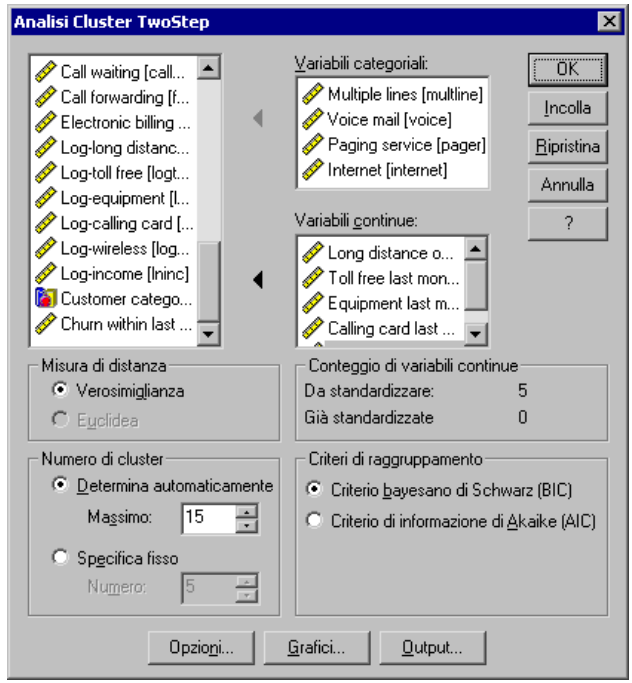
- **Gestione di variabili categoriali e continue.** Se le variabili sono indipendenti, è possibile applicare una distribuzione normale multinomiale congiunta alle variabili categoriali e continue.
- **Selezione automatica del numero di cluster.** Mediante il confronto tra i valori dei criteri di scelta di modello appartenenti a diverse soluzioni di raggruppamento, la procedura è in grado di determinare automaticamente il numero ottimale di cluster.
- **Scalabilità.** Mediante la creazione di un albero delle caratteristiche dei cluster (CF) che fornisce un riepilogo dei record, l'algoritmo TwoStep consente di analizzare file di dati di grandi dimensioni.

Esempio. I produttori di articoli al dettaglio e prodotti per i consumatori applicano regolarmente tecniche di raggruppamento ai dati relativi alle abitudini di acquisto dei propri clienti, al sesso, all'età, al livello di reddito e così via. In questo modo adattano le strategie di sviluppo del prodotto e di mercato ad ogni gruppo di consumatori al fine di aumentare le vendite e accrescere la fedeltà alla marca.

Statistiche. La procedura fornisce criteri di informazione (AIC o BIC) in base ai numeri di cluster nella soluzione, frequenze di cluster per il raggruppamento finale e statistiche descrittive in base al cluster per il raggruppamento finale.

Grafici. La procedura fornisce grafici a barre e grafici a torta per le frequenze dei cluster e grafici relativi all'importanza delle variabili.

Figura 32-1
Finestra di dialogo Analisi Cluster TwoStep



Misura di distanza. Questa selezione determina la modalità di calcolo della similarità tra due cluster.

- **Verosimiglianza.** La misura di verosimiglianza applica una distribuzione per probabilità alle variabili. Si suppone che le variabili continue vengano distribuite normalmente, mentre le variabili categoriali in base al modello multinomiale. Si suppone che tutte le variabili siano indipendenti.
- **Euclidea.** La misura euclidea è la distanza in "linea retta" tra due cluster. Può essere utilizzata solo quando tutte le variabili sono continue.

Numero di cluster. Questa selezione consente di specificare la modalità di definizione del numero dei cluster.

- **Determina automaticamente.** La procedura determina automaticamente il numero di cluster ottimale, mediante i criteri specificati nel gruppo Criteri di raggruppamento. È inoltre possibile immettere un numero intero positivo

per definire il numero di cluster massimo che la procedura dovrà prendere in considerazione.

- **Specifica fisso.** Consente di definire un numero fisso di cluster nella soluzione. Immettere un numero intero positivo.

Conteggio di variabili continue. Questo gruppo fornisce un riepilogo delle opzioni di standardizzazione relative alle variabili continue specificate nella finestra di dialogo Opzioni. Per ulteriori informazioni, vedere "Opzioni di Analisi Cluster TwoStep" a pag. 511.

Criteri di raggruppamento. Questa selezione determina la modalità di definizione del numero dei cluster mediante l'algoritmo di raggruppamento automatico. È possibile specificare il modello Criterio bayesiano di Schwarz (BIC, Bayesian Information Criterion) o il modello Criterio di informazione di Akaike (AIC, Akaike Information Criterion).

Dati. Questa procedura può essere utilizzata sia con le variabili continue sia con le variabili categoriali. I casi rappresentano gli oggetti da raggruppare e le variabili corrispondono agli attributi in base ai quali viene eseguito il raggruppamento.

Ordine dei casi. La funzione Albero delle caratteristiche cluster e la soluzione finale possono dipendere dall'ordine dei casi. Per ridurre al minimo gli effetti dell'ordine, disporre i casi in ordine casuale. Può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica. Nei casi in cui questa operazione è complessa a causa delle dimensioni eccessive dei file, è possibile effettuare più operazioni con un campione di casi disposti in ordini casuali diversi.

Assunzioni. La misura di distanza della verosimiglianza assume che le variabili nel modello di cluster siano indipendenti. A ogni variabile continua inoltre si suppone inoltre che sia associata una distribuzione normale o gaussiana, mentre a ogni variabile categoriale una distribuzione multinomiale. La verifica empirica interna indica che la procedura è piuttosto robusta rispetto alle violazioni sia delle assunzioni di indipendenza sia delle assunzioni di distribuzione, ma è consigliabile verificare fino a che punto tali assunzioni vengano soddisfatte.

Utilizzare la proceduraCorrelazioni bivariate per verificare l'indipendenza di due variabili continue. Utilizzare la proceduraTavole di contingenza per verificare l'indipendenza di due variabili categoriali. Utilizzare la proceduraMedie per verificare l'indipendenza tra una variabile continua e una variabile categoriale. Utilizzare la proceduraEsplora per verificare la normalità di una variabile continua. Utilizzare la

proceduraTest Chi-quadrato per verificare se per una variabile categoriale è stata specificata una distribuzione multinomiale.

Per ottenere un'analisi Cluster TwoStep

- Dai menu, scegliere:

  Analizza
  Classificazione
  Cluster TwoStep...

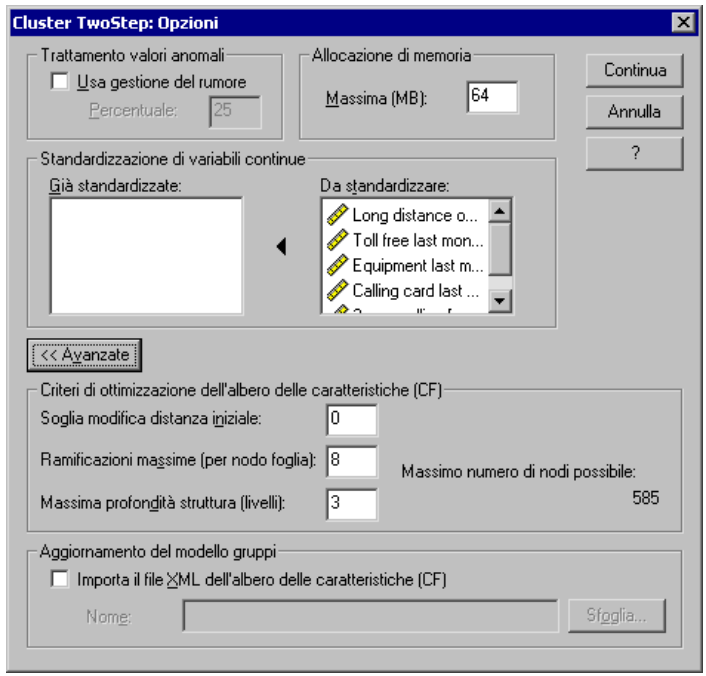
- Selezionare una o più variabili categoriali o continue.

Se lo si desidera, è possibile:

- Modificare i criteri in base ai quali sono stati creati i cluster.
- Selezionare le impostazioni di gestione del rumore, allocazione di memoria, standardizzazione delle variabili e input del modello di cluster.
- Richiedere tabelle e grafici facoltativi.
- Salvare i risultati del modello nel file di lavoro o in un file XML esterno.

Opzioni di Analisi Cluster TwoStep

Figura 32-2
Finestra di dialogo Analisi Cluster TwoStep: Opzioni



Trattamento valori anomali. Questo gruppo consente di trattare i valori anomali soprattutto durante il raggruppamento, se l'albero delle caratteristiche dei cluster risulta pieno. L'albero delle caratteristiche dei cluster è pieno quando non è più in grado di accettare casi in un nodo foglia e nessun nodo foglia può essere suddiviso.

- Se si seleziona la gestione del rumore e l'albero delle caratteristiche dei cluster risulta pieno, sarà possibile ampliarlo posizionando i casi mal distribuiti in più foglie all'interno di una foglia specifica per il "rumore". Una foglia contiene casi mal distribuiti quando il numero dei casi è inferiore alla percentuale specificata per la dimensione massima della foglia. Dopo aver ampliato la struttura, i valori

anomali vengono inseriti nell'albero delle caratteristiche dei cluster, se possibile. Altrimenti, vengono eliminati.

- Se non si seleziona la gestione del rumore e l'albero delle caratteristiche dei cluster risulta pieno, sarà possibile ampliarlo utilizzando una soglia di modifica della distanza più elevata. Dopo il raggruppamento finale, i valori non assegnati a un cluster vengono definiti valori anomali. Al cluster dei valori anomali viene assegnato il numero di identificazione –1 e non viene incluso nel conteggio dei cluster.

Allocazione di memoria. Questo gruppo consente di specificare in megabyte (MB) la quantità massima di memoria che l'algoritmo del cluster può utilizzare. Se questa quantità massima viene superata, la procedura utilizzerà il disco per memorizzare le informazioni che non è possibile inserire nella memoria. Specificare un numero maggiore o uguale a 4.

- Per informazioni sul valore massimo per il sistema, rivolgersi all'amministratore del sistema.
- La ricerca dei cluster corretti o desiderati mediante l'algoritmo potrebbe non riuscire, se il valore è troppo basso.

Standardizzazione delle variabili. L'algoritmo di raggruppamento funziona con variabili continue standardizzate. Qualsiasi variabile non standardizzata deve essere impostata come "Da standardizzare". Per risparmiare tempo e calcoli, è possibile impostare le variabili continue già standardizzate come "Già standardizzate".

Opzioni avanzate

Criteri di ottimizzazione dell'albero delle caratteristiche (CF). Le seguenti impostazioni dell'algoritmo di raggruppamento riguardano in modo specifico l'albero delle caratteristiche dei cluster e devono essere modificate con cautela:

- **Soglia modifica distanza iniziale.** Si tratta della soglia iniziale utilizzata per ampliare l'albero delle caratteristiche dei cluster. Se dopo l'inserimento di un determinato caso in una foglia dell'albero delle caratteristiche dei cluster, la distanza risulta inferiore alla soglia, la foglia non viene suddivisa. Se la distanza supera la soglia, la foglia può essere suddivisa.
- **Ramificazioni massime (per nodo foglia).** Il numero massimo di nodi figlio per un nodo foglia.

- **Massima profondità struttura (livelli).** Il numero massimo di livelli dell'albero delle caratteristiche dei cluster.
- **Massimo numero di nodi possibile.** Indica il numero massimo dei nodi dell'albero delle caratteristiche dei cluster che può essere potenzialmente generato dalla procedura, in base alla funzione $(b^{d+1} - 1) / (b - 1)$, dove b sono le ramificazioni massime e d la profondità massima. Tenere presente che un albero delle caratteristiche dei cluster di dimensioni eccessive può rappresentare un peso considerevole per le risorse del sistema e compromettere quindi le prestazioni della procedura. Ogni nodo richiede un minimo di 16 byte.

Aggiornamento del modello gruppi. Questo gruppo consente di importare e aggiornare un modello di cluster generato da un'analisi precedente. Il file di input contiene l'albero delle caratteristiche dei cluster in formato XML. Il modello viene quindi aggiornato con i dati nel file attivo. È necessario selezionare i nomi delle variabili nella finestra di dialogo principale in base allo stesso ordine dell'analisi precedente. Il file XML non viene modificato, a meno che le nuove informazioni relative al modello non vengano inserite nello stesso file. Per ulteriori informazioni, vedere "Output di Analisi Cluster TwoStep" a pag. 515.

Se viene selezionato l'aggiornamento del modello dei cluster, verranno utilizzate le opzioni per la generazione dell'albero delle caratteristiche dei cluster specificate per il modello originale. Vengono quindi utilizzate le impostazioni del modello salvato relative a misura di distanza, gestione del rumore, allocazione di memoria e ottimizzazione dell'albero delle caratteristiche dei cluster, mentre qualsiasi nuova impostazione specificata nelle finestre di dialogo viene ignorata.

Nota: durante l'aggiornamento di un modello di cluster, la procedura si basa sul presupposto che nessuno dei casi selezionati nel file dati attivo sia stato utilizzato per creare il modello di cluster originale. La procedura si basa inoltre sul presupposto che i casi utilizzati nell'aggiornamento del modello di cluster provengono dalla stessa popolazione di casi utilizzati per creare il modello originale, le medie e le varianze delle variabili continue e i livelli delle variabili categoriali devono quindi essere le stesse per i due insiemi di casi. Se gli insiemi di casi precedenti e correnti provengono da una popolazione eterogenea, è necessario eseguire la procedura Analisi Cluster TwoStep negli insiemi di casi combinati per ottenere risultati ottimali.

Grafici di Analisi Cluster TwoStep

Figura 32-3
Finestra di dialogo Cluster TwoStep: Grafici

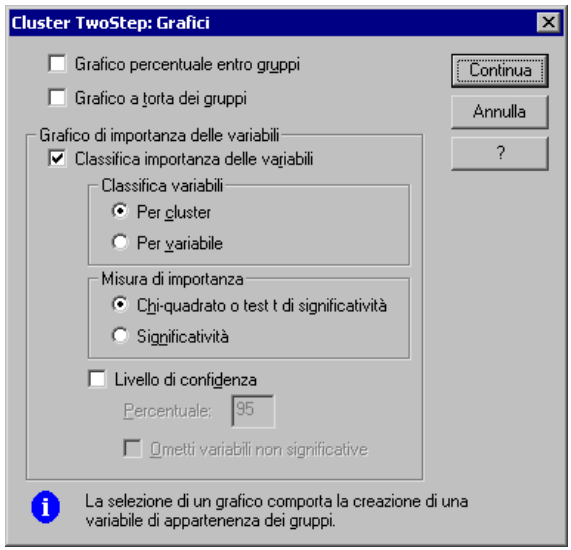


Grafico percentuale entro gruppi. Visualizza i grafici della variazione di ogni variabile all'interno di un cluster. Per ogni variabile categoriale viene creato un grafico a barre raggruppato, in cui viene illustrata la frequenza delle categorie in base all'ID cluster. Per ogni variabile continua viene invece creato un grafico degli errori, in cui vengono riportati gli errori in base all'ID cluster.

Grafico a torta dei gruppi. Visualizza un grafico a torta con le percentuali e i conteggi delle osservazioni all'interno di ogni cluster.

Grafico di importanza delle variabili. Visualizza diversi grafici in cui viene illustrata l'importanza di ogni variabile all'interno di ogni cluster. L'output viene ordinato in base al livello di importanza di ogni variabile.

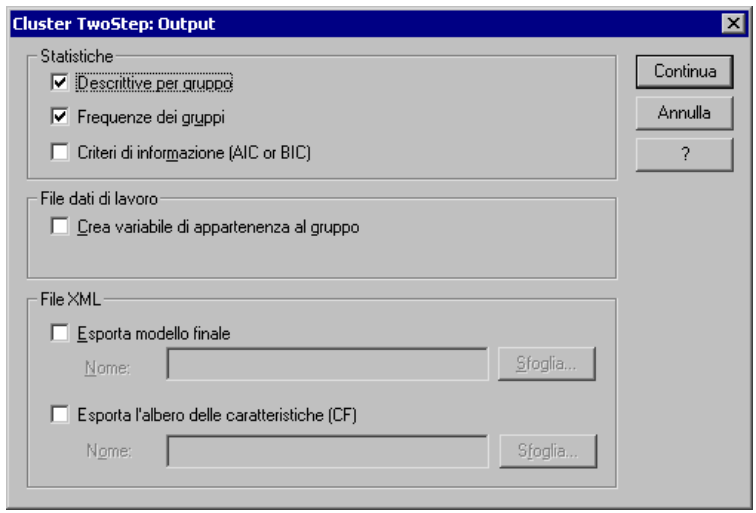
- **Classifica variabili.** Questa opzione consente di definire se i grafici verranno creati per ogni cluster (Per cluster) o per ogni variabile (Per variabile).
- **Misura di importanza.** Questa opzione consente di selezionare la misura di importanza della variabile che verrà utilizzata per creare il grafico. Chi-quadrato o test t di significatività segnala una statistica chi-quadrato di Pearson come importanza della variabile categoriale e una statistica *t* come importanza della

variabile continua. Significatività segnala il valore *p* meno uno per il test di uguaglianza delle medie di una variabile continua e la frequenza attesa per l'insieme di dati globale di una variabile categoriale.

- **Livello di confidenza.** Questa opzione consente di impostare il livello di confidenza per il test di uguaglianza della distribuzione di una variabile all'interno di un cluster rispetto alla distribuzione globale della variabile. Specificare un numero minore di 100 e maggiore o uguale a 50. Il valore del livello di confidenza è rappresentato da una riga verticale nei grafici di importanza delle variabili, se i grafici vengono creati in base alla variabile o se la misura di significatività è riportata nel grafico.
- **Ometti variabili non significative.** Le variabili non significative in base al livello di confidenza specificato non vengono visualizzate nei grafici di importanza delle variabili.

Output di Analisi Cluster TwoStep

Figura 32-4
Finestra di dialogo Analisi Cluster TwoStep: Output



Statistiche. Questo gruppo fornisce le opzioni per la visualizzazione delle tabelle dei risultati dei raggruppamenti. Le statistiche descrittive e le frequenze dei cluster vengono fornite per il modello di cluster finale, mentre la tabella dei criteri delle informazioni contiene i risultati di un intervallo di soluzioni di cluster.

- **Descrittive per gruppo.** Visualizza due tabelle che descrivono le variabili presenti in ogni cluster. In una tabella, le medie e le deviazioni standard vengono segnalate per le variabili continue in base al cluster. L'altra tabella riporta le frequenze delle variabili categoriali per cluster.
- **Frequenze dei gruppi.** Visualizza una tabella che riporta il numero di osservazioni in ogni cluster.
- **Criteri di informazione (AIC or BIC).** Visualizza una tabella che include i valori di AIC o BIC, in base ai criteri scelti nella finestra di dialogo principale, per i diversi numeri di cluster. Questa tabella viene fornita solo quando il numero di cluster viene determinato in modo automatico. Se il numero di cluster è fisso, questa impostazione viene ignorata e la tabella non viene fornita.

File di dati attivo. Questo gruppo consente di salvare le variabili all'interno del file dati attivo.

- **Crea variabile di appartenenza al gruppo.** Questa variabile contiene un numero di identificazione del cluster per ogni caso. Il nome di questa variabile è *tsc_n*, dove *n* è un numero intero positivo che indica l'ordinale dell'operazione di salvataggio del file dati attivo eseguita mediante questa procedura in una determinata sessione.

File XML. Il modello di cluster finale e l'albero delle caratteristiche dei cluster rappresentano due tipi di file di output che possono essere esportati in formato XML.

- **Esporta modello finale.** Il modello di cluster finale viene esportato nel file specificato in formato XML (PMML). SmartScore e la versione del server di SPSS (un prodotto separato) possono utilizzare questo file di modello per applicare le informazioni del modello ad altri file dati a scopo di assegnazione dei punteggi.
- **Esporta l'albero delle caratteristiche (CF).** Questa opzione consente di salvare lo stato corrente dell'albero delle caratteristiche dei cluster e aggiornarlo in un secondo momento utilizzando dati più aggiornati.

Capitolo
33

Cluster gerarchica

Questa procedura consente di identificare gruppi di casi relativamente omogenei in base alle caratteristiche selezionate, utilizzando un algoritmo che inizia con ciascun caso (o variabile) in un cluster distinto e che combina i cluster fino a quando ne rimane solo uno. È possibile analizzare le variabili semplici oppure scegliere una delle trasformazioni di standardizzazione disponibili. Le misure di similarità e dissimilarità vengono generate dalla procedura Distanze. A ciascun livello verranno visualizzate statistiche in base alle quali selezionare la soluzione migliore.

Esempio. Esistono gruppi di trasmissioni televisive identificabili che attraggono tipi di audience analoghi all'interno di ciascun gruppo? Utilizzando la cluster gerarchica è possibile raggruppare le trasmissioni televisive (casi) in gruppi omogenei in base alle caratteristiche degli spettatori. Questo metodo può essere utilizzato per identificare i segmenti di mercato. In alternativa, è possibile raggruppare le città (casi) in gruppi omogenei in modo che da poter selezionare città con caratteristiche confrontabili per verificare diverse strategie di mercato.

Statistiche. Programma di agglomerazione, matrice delle distanze (o similarità) e cluster di appartenenza per un'unica soluzione o una serie di soluzioni. Grafici: dendrogrammi e grafici a stalattite

Dati. Le variabili possono essere quantitative, binarie o dati di conteggio. Lo scaling delle variabili è molto importante in quanto le differenze di scaling possono influire sulle soluzioni dei cluster. Se lo scaling delle variabili presenta differenze notevoli (ad esempio, una variabile viene misurata in dollari e l'altra in anni), è consigliabile standardizzarle. Ciò può essere effettuato in modo automatico mediante la procedura Cluster gerarchica.

Ordine dei casi. Se le distanze assegnate o le similarità sono presenti nei dati iniziali o nei cluster aggiornati durante l'unione, la soluzione del cluster risultante può essere influenzata dall'ordine dei casi del file. Può essere utile ottenere più soluzioni diverse

con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica.

Assunzioni. Le misure di dissimilarità o di similarità utilizzate devono essere idonee per i dati analizzati. Per ulteriori informazioni sulla scelta delle misure di dissimilarità e similarità, vedere la procedura Distanze. È inoltre necessario includere nell'analisi tutte le variabili significative. L'omissione di variabili importanti può portare a soluzioni improprie. Poiché la cluster gerarchica rappresenta un metodo esplorativo, i risultati devono essere considerati provvisori finché non vengano confermati da un campione indipendente.

Figura 33-1
Output della cluster gerarchica

Programma di agglomerazione

		Cluster accorpati			Stadio di formazione del cluster		
		Cluster 1	Cluster 2	Coefficienti	Cluster 1	Cluster 2	Prossimo stadio
Stadio	1	11	12	,112	0	0	2
	2	6	11	,132	0	1	4
	3	7	9	,185	0	0	5
	4	6	8	,227	2	0	7
	5	7	10	,274	3	0	7
	6	1	3	,423	0	0	10
	7	6	7	,438	4	5	14
	8	13	14	,484	0	0	15
	9	2	5	,547	0	0	11
	10	1	4	,691	6	0	11
	11	1	2	1,023	10	9	13
	12	15	16	1,370	0	0	13
	13	1	15	1,716	11	12	14
	14	1	6	2,642	13	7	15
	15	1	13	4,772	14	8	0

Cluster di appartenenza

		Etichetta	4 Cluster	3 Cluster	2 Cluster
Caso	1	Argentina	1	1	1
	2	Brazil	1	1	1
	3	Chile	1	1	1
	4	Domincan R.	1	1	1
	5	Indonesia	1	1	1
	6	Austria	2	2	1
	7	Canada	2	2	1
	8	Denmark	2	2	1
	9	Italy	2	2	1
	10	Japan	2	2	1
	11	Norway	2	2	1
	12	Switzerland	2	2	1
	13	Bangladesh	3	3	2
	14	India	3	3	2
	15	Bolivia	4	1	1
	16	Paraguay	4	1	1

Grafico a stalattite verticale

Numero di cluster	Caso																													
	14 India	14	13 Banglade	13	10 Japan	10	9 Italy	9	7 Canada	7	8 Denmark	8	12 Switzerlar	12	11 Norway	11	6 Austria	6	16 Paraguay	16	15 Bolivia	15	5 Indonesia	5	2 Brazil	2	4 Domincan	4	3 Chile	3	1 Argentina
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
4	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X	X	X		X	X	X	X	X	X	X	X	X
5	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X	X	X	X	X	X	X	X	X
6	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X	X	X		X	X	X	X	X
7	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X	X	X		X		X	X	X
8	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X		X		X		X	X	X
9	X		X		X	X	X	X	X	X	X	X	X	X	X	X	X		X		X		X		X		X		X	X	X
10	X		X		X	X	X	X	X		X	X	X	X	X	X	X		X		X		X		X		X		X	X	X
11	X		X		X	X	X	X	X		X	X	X	X	X	X	X		X		X		X		X		X		X		X
12	X		X		X		X	X	X		X	X	X	X	X	X	X		X		X		X		X		X		X		X
13	X		X		X		X	X	X		X		X	X	X	X	X		X		X		X		X		X		X		X
14	X		X		X		X		X		X		X	X	X	X	X		X		X		X		X		X		X		X
15	X		X		X		X		X		X		X	X	X		X		X		X		X		X		X		X		X

```
* * * * * * * * * * * * * * CLUSTER GERARCHICA* * * * * * * * * * * * * * * *

   Dendrogramma con legame medio fra i gruppi

                Cluster accorpati con distanza riscalata

 CASO             0         5        10        15        20        25
Etichetta Numero  +---------+---------+---------+---------+---------+

LIFEEXPF     2    ßÛÛÛÛÛÙ
BABYMORT     5    ßÛÛÙ  ßÛÛÙ
LITERACY     3    ßÛÛÛÛÛÛÛÛÙ  ßÛÛÛÛÛÛÛÛÛÛÙ
BIRTH_RT     6    ßÛÛÛÛÛÛÛÛÛÛÙ  ßÛÛÛÛÛÛÛÛÛÛÛÛÛÛÙ
FERTILTY    10    ßÛÛÛÛÙ
URBAN        1    ßÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÙ  ßÛÛÛÛÛÛÛÛÙ
LOG_GDP      8    ßÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÙ
POP_INCR     4    ßÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÙ
B_TO_D       9    ßÛÛÛÛÛÛÛÛÙ  ßÛÛÛÛÛÙ
DEATH_RT     7    ßÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÙ
LOG_POP     11    ßÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÛÙ
```

Per ottenere una cluster gerarchica

- Dai menu, scegliere:

Analizza
  Classificazione
    Cluster gerarchica...

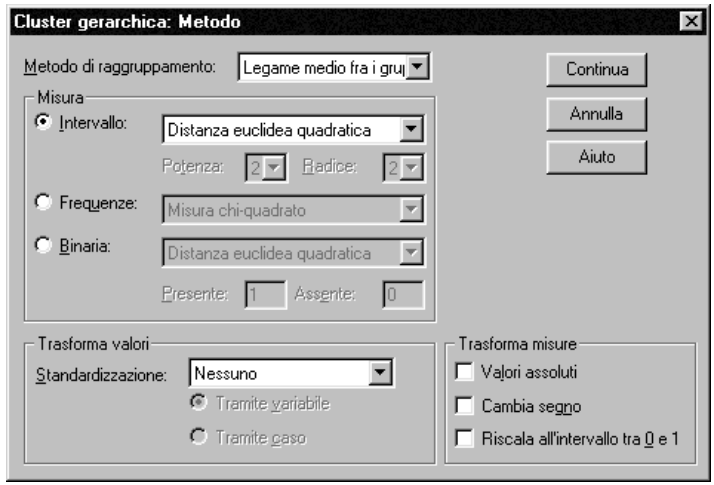
Figura 33-2
Finestra di dialogo Cluster gerarchica

- Per raggruppare i casi in cluster è necessario selezionare almeno una variabile numerica. Per raggruppare le variabili in cluster è necessario selezionare almeno tre variabili numeriche.

È inoltre possibile selezionare una variabile di identificazione per etichettare i casi.

Cluster gerarchica: Metodo

Figura 33-3
Finestra di dialogo Cluster gerarchica: Metodo



Metodo di raggruppamento. Le alternative disponibili sono: Legame medio fra i gruppi, Legame medio entro gruppi, Del vicino più vicino, Del vicino più lontano, Centroide, Mediana e Ward.

Misura. Consente di specificare la misura di similarità o dissimilarità da utilizzare per il raggruppamento. Selezionare il tipo di dati e la misura di similarità o dissimilarità desiderata:

- **Intervallo.** Le alternative disponibili sono: Distanza euclidea, Distanza euclidea quadratica, Coseno, Correlazione di Pearson, Chebychev, City-Block, Minkowski e Personalizzato.
- **Conteggi.** Le alternative disponibili sono: Misura chi-quadrato e Misura phi-quadrato.
- **Binaria.** Le alternative disponibili sono: Distanza euclidea, Distanza euclidea quadratica, Differenza di dimensione, Differenza di modello, Varianza, Dispersione, Forma, Corrispondenza semplice, Correlazione phi a 4 punti, Lambda, *D* di Anderberg, Dice, Hamann, Jaccard, Kulczynski 1, Kulczynski 2, Lance e Williams, Ochiai, Rogers e Tanimoto, Russel e Rao, Sokal e Sneath 1,

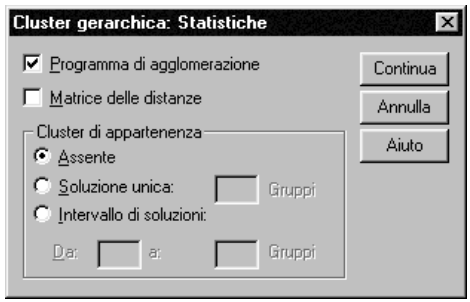
Sokal e Sneath 2, Sokal e Sneath 3, Sokal e Sneath 4, Sokal e Sneath 5, *Y* di Yule e *Q*di Yule.

Trasforma valori. Consente di standardizzare i valori dei dati per casi o valori prima di calcolare le similarità (non disponibile per i dati binari). I metodi di standardizzazione disponibili sono: punteggi *z*, intervallo da –1 a 1, intervallo da 0 a 1, ampiezza massima di 1, media di 1 e deviazione standard di 1.

Trasforma misure. Consente di trasformare i valori generati dalla misura di distanza. Questi verranno applicati dopo il calcolo della misura di distanza. Le alternative disponibili sono: Valori assoluti, Cambia segno e Riscala all'intervallo fra 0 e 1.

Cluster gerarchica: Statistiche

Figura 33-4
Finestra di dialogo Cluster gerarchica: Statistiche



Programma di agglomerazione. Consente di visualizzare i casi o i cluster accorpati ad ogni stadio, le distanze tra i casi o i cluster da accorpare e l'ultimo livello di cluster in cui un caso (o una variabile) è stato accorpato al cluster.

Matrice delle distanze. Fornisce le distanze o le similarità tra gli elementi.

Cluster di appartenenza. Viene visualizzato il cluster a cui viene assegnato ciascun caso a uno o più stadi della combinazione dei cluster. Le opzioni disponibili sono Soluzione unica e Intervallo di soluzioni.

Cluster gerarchica: Grafici

Figura 33-5
Finestra di dialogo Cluster gerarchica: Grafici

Dendrogramma. Visualizza un **dendrogramma**. Utilizzando i dendrogrammi è possibile valutare la coesione dei cluster formati ed ottenere informazioni sul numero di cluster che è opportuno tenere.

A stalattite. Visualizza un **grafico a stalattite**, che comprende tutti i cluster o un intervallo specifico di cluster. Nei grafici a stalattite vengono visualizzate informazioni sulle modalità con cui i casi vengono combinati in cluster ad ogni iterazione dell'analisi. Specificando l'orientamento desiderato è possibile selezionare un grafico verticale o orizzontale.

Cluster gerarchica: Salva nuove variabili

Figura 33-6
Finestra di dialogo Cluster gerarchica: Salva

Cluster di appartenenza. Consente di salvare i cluster di appartenenza per una soluzione unica o per un intervallo di soluzioni. Le variabili salvate possono essere utilizzate in analisi successive per valutare altre differenze tra i gruppi.

Funzioni aggiuntive della sintassi del comando CLUSTER

La procedura Cluster gerarchica usa la sintassi del comando `CLUSTER`. Il linguaggio a comandi SPSS permette anche di:

- Usare più metodi di raggruppamento in una singola analisi.
- Leggere ed analizzare una matrice di prossimità.
- Scrivere una matrice di prossimità sul disco per analizzarla in seguito.
- Specificare i valori per la potenza e la radice nella misura della distanza personalizzata (potenza).
- Specificare i nomi delle variabili salvate.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
34
Cluster con metodo delle K-medie

Questa procedura consente di identificare gruppi di casi relativamente omogenei in base alle caratteristiche selezionate, utilizzando un algoritmo in grado di gestire un elevato numero di casi. Tale algoritmo, tuttavia, richiede l'indicazione del numero di cluster. È possibile specificare i centri iniziali del cluster, se si conosce questa informazione. È possibile selezionare uno dei due metodi disponibili per la classificazione dei casi, ovvero l'aggiornamento iterativo dei centri cluster oppure la semplice classificazione. È possibile salvare l'appartenenza al cluster, le informazioni sulla distanza e i centri del cluster finali. È inoltre possibile specificare una variabile i cui valori possono essere utilizzati per etichettare l'output caso per caso. i può inoltre richiedere l'analisi delle statistiche *F* di varianza. Se da un lato queste statistiche sono opportunistiche, ovvero vengono eseguiti tentativi di raggruppamenti che presentino differenze, le corrispondenti dimensioni relative forniscono informazioni sul contributo apportato da ciascuna variabile alla separazione dei gruppi.

Esempio. Quali sono i gruppi di show televisivi che attraggono un pubblico analogo all'interno di ciascun gruppo? Il metodo cluster *k*-medie consente di raggruppare gli show televisivi (casi) in*k* gruppi omogenei in base alle caratteristiche degli spettatori. Questo processo può essere utilizzato per identificare i segmenti di mercato. In alternativa, è possibile raggruppare le città (casi) in gruppi omogenei in modo che da poter selezionare città con caratteristiche confrontabili per verificare diverse strategie di mercato.

Statistiche. Soluzione completa: centri iniziali del cluster, tabella ANOVA. Per ciascun caso: informazioni sui cluster, distanza dal centro del cluster.

Dati. Le variabili devono essere quantitative a livello di intervallo o di rapporto. Se le variabili sono binarie o conteggi, utilizzare la procedura Cluster gerarchica.

Ordine dei casi e dei centri di cluster iniziale. L'algoritmo predefinito per la scelta dei centri di cluster iniziali varia a seconda dell'ordine dei casi. L'opzioneUsa medie mobili della finestra di dialogo Iterazioni rende la soluzione risultante potenzialmente

dipendente dall'ordine dei casi, indipendentemente dai centri di cluster scelti inizialmente. Se si utilizza uno di questi metodi, può essere utile ottenere più soluzioni diverse con casi disposti in ordini casuali diversi per verificare la stabilità di una soluzione specifica. Per evitare problemi con l'ordine dei casi, è consigliabile specificare i centri di cluster iniziali ed evitare di usare l'opzione Usa medie mobili. Tuttavia, l'ordinamento dei centri di cluster iniziali può influire sulla soluzione se esistono distanze assegnate dai casi ai centri di cluster. Per valutare la stabilità di una soluzione, è possibile confrontare i risultati delle analisi con diverse permutazioni dei valori dei centri iniziali.

Assunzioni. Le distanze vengono calcolate utilizzando la distanza euclidea semplice. Se si desidera utilizzare un'altra misura di distanza o di similarità, utilizzare la procedura Cluster gerarchica. La scalatura delle variabili è un'operazione che deve essere effettuata con molta attenzione. Se le variabili vengono misurate con scale diverse (ad esempio se una variabile è espressa in dollari e un'altra è espressa in anni), i risultati possono essere fuorvianti. In questi casi è consigliabile standardizzare le variabili prima di procedere con l'analisi cluster *k*medie (utilizzando la procedura Descrittive). Questa procedura presume che sia stato selezionato il numero esatto di cluster e che siano state incluse tutte le variabili rilevanti. Se è stato selezionato un numero di cluster inesatto o sono state omesse variabili importanti, i risultati possono essere inattendibili.

Figura 34-1
Output della procedura Cluster K-medie

Centri iniziali dei cluster

	Cluster			
	1	2	3	4
ZURBAN	-1,88606	-1,54314	1,45741	,55724
ZLIFEEXP	-3,52581	-1,69358	,62725	,99370
ZLITERAC	-2,89320	-1,65146	-,51770	,88601
ZPOP_INC	,93737	,16291	3,03701	-1,12785
ZBABYMOR	4,16813	1,38422	-,69589	-,88983
ZBIRTH_R	2,68796	,42699	,33278	-1,08033
ZDEATH_R	4,41517	,63185	-1,89037	,63185
ZLOG_GDP	-1,99641	-1,78455	,53091	1,22118
ZB_TO_D	-,52182	-,31333	4,40082	-,99285
ZFERTILT	2,24070	,75481	,46008	-,76793
ZLOG_POP	,24626	2,65246	-1,29624	-,74406

Cronologia iterazioni

	Modifiche ai centri dei cluster			
Iterazione	1	2	3	4
1	1.932	2.724	1.596	3.343
2	.000	.471	.314	.466
3	.861	.414	.195	.172
4	.604	.337	.150	.000
5	.000	.253	.167	.237
6	.000	.199	.071	.287
7	.623	.160	.000	.000
8	.000	.084	.074	.000
9	.000	.080	.077	.000
10	.000	.097	.000	.185

Centri dei cluster finali

	Cluster			
	1	2	3	4
ZURBANA	-1.70745	-.30863	.62767	.16816
ZSPVITA	-2.52826	-.15939	.80611	-.28417
ZALFAB	-2.30833	.13880	.73368	-.81671
ZINCRPOP	.59747	.13400	-.95175	1.45301
ZMORTINF	2.43210	.22286	-.80817	.25622
ZTASNAT	1.52607	.12929	-.99285	1.13716
ZTASMOR	2.10314	-.44640	.31319	-.71414
ZLOGPIL	-1.77704	-.58745	.94249	-.16871
ZNATSUMO	-.29856	.19154	-.84758	1.45251
ZFERTIL	1.51003	-.12150	-.87669	1.27010
ZLOGPOP	.83475	.34577	-.22199	-.49499

Distanze tra i centri dei cluster finali

Cluster	1	2	3	4
1		5.627	7.924	5.640
2	5.627		3.249	2.897
3	7.924	3.249		5.246
4	5.640	2.897	5.246	

ANOVA

	Cluster		Errore			
	Media dei quadrati	df	Media dei quadrati	df	F	Sig.
ZURBANA	10.409	3	.541	68	19.234	.000
ZSPVITA	19.410	3	.210	68	92.614	.000
ZALFAB	18.731	3	.229	68	81.655	.000
ZINCRPOP	18.464	3	.219	68	84.428	.000
ZMORTINF	18.621	3	.239	68	77.859	.000
ZTASNAT	19.599	3	.167	68	117.339	.000
ZTASMOR	13.628	3	.444	68	30.676	.000
ZLOGPIL	17.599	3	.287	68	61.313	.000
ZNATSUMO	16.316	3	.288	68	56.682	.000
ZFERTIL	18.829	3	.168	68	112.273	.000
ZLOGPOP	3.907	3	.877	68	4.457	.006

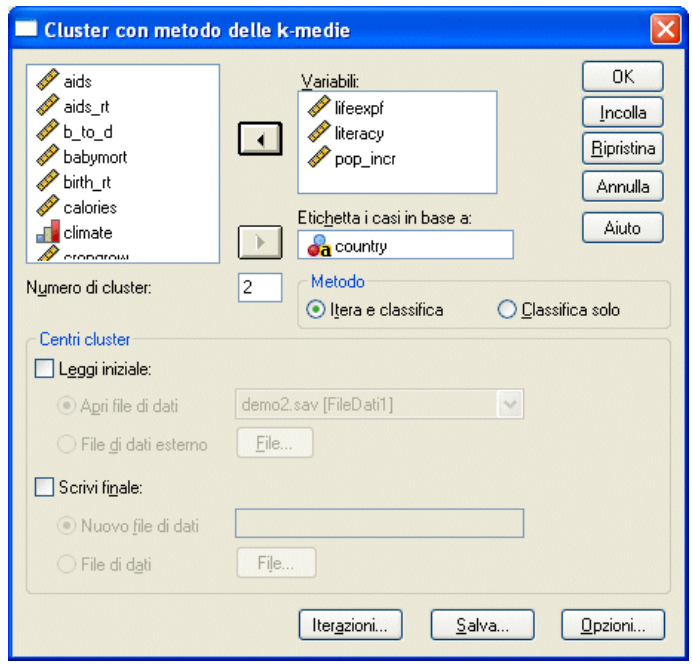
I test F devono essere utilizzati solo per motivi descrittivi poiché i cluster sono stati scelti per ottimizzare le differenze tra i casi in diversi cluster. I livelli di significatività osservati non sono perciò corretti e non possono quindi essere interpretati come test dell'ipotesi che le medie dei cluster siano uguali.

Per ottenere un'analisi cluster K-medie

- Dai menu, scegliere:

  Analizza
  Classificazione
  Cluster K-medie...

Figura 34-2
Finestra di dialogo Cluster K-medie



- Selezionare le variabili da utilizzare nell'analisi cluster.
- Specificare il numero di cluster. Il numero di cluster specificato deve essere almeno di 2 e non deve essere maggiore al numero di casi del file dati.
- Selezionare il metodo Itera e classifica oppure il metodo Classifica soltanto.
- In alternativa, selezionare una variabile di identificazione per etichettare i casi.

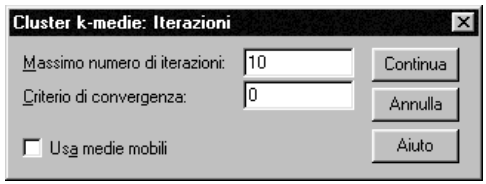
Efficienza dell'analisi cluster K-medie

Il comando Cluster *k*-medie è efficace principalmente in quanto non calcola le distanze tra tutte le coppie di casi, a differenza di numerosi algoritmi di raggruppamento, ad esempio quello utilizzato dal comando per la Cluster gerarchica di SPSS.

Per ottenere la massima efficienza, creare un campione di casi e utilizzare il metodo Itera e classifica per determinare i centri cluster. Selezionare Scrivi valori finali su file. Quindi, ripristinare tutto il file di dati e selezionare Classifica soltanto come metodo e selezionare Leggi valori iniziali per classificare tutto il file utilizzando i centri valutati per il campione. È possibile leggere o scrivere da un file o file di dati. I file di dati possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali alla fine della sessione. I nomi dei file di dati devono essere conformi alle regole dei nomi delle variabili di SPSS. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.

Cluster K-medie: Iterazioni

Figura 34-3
Finestra di dialogo Cluster K-medie: Iterazioni



Nota: queste opzioni sono disponibili solo se si seleziona il metodo Itera e classifica nella finestra di dialogo Cluster con metodo delle K-medie.

Massimo numero di iterazioni. Consente di impostare il numero massimo di iterazioni per l'algoritmo *k*-medie. Le iterazioni si interromperanno al numero impostato, anche se il criterio di convergenza non viene soddisfatto. Il numero deve essere compreso tra 1 e 999.

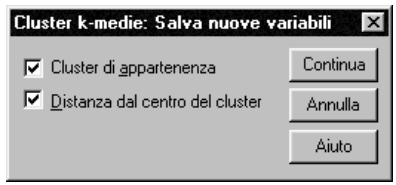
Per riprodurre l'algoritmo utilizzato dal comando Quick Cluster delle versioni di SPSS precedenti alla 5.0, impostare l'opzione Massimo numero di iterazioni su 1.

Criterio di convergenza. Determina il termine dell'iterazione. Rappresenta una proporzione della distanza minima fra i centri iniziali del cluster in modo che sia maggiore di 0 e minore di 1. Se, ad esempio, il criterio è 0,02, il processo di iterazione terminerà quando un'iterazione completa non è in grado di spostare i centri cluster di una distanza maggiore del 2% della distanza minima fra i centri iniziali del cluster.

Usa medie mobili. Consente di richiedere l'aggiornamento dei centri cluster in seguito all'assegnazione di ciascun caso Se non viene selezionata questa opzione, i nuovi centri del cluster verranno calcolati quando tutti i casi saranno stati assegnati.

Cluster K-medie: Salva

Figura 34-4
Finestra di dialogo Cluster K-medie: Salva nuove variabili



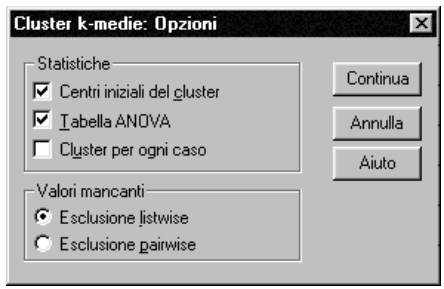
È possibile salvare informazioni sulla soluzione come nuove variabili da utilizzare in analisi successive:

Cluster di appartenenza. Consente di creare una nuova variabile che indica l'appartenenza finale al cluster di ciascun caso. I valori della nuova variabile sono compresi tra 1 e il numero di cluster.

Distanza dal centro. Consente di creare una nuova variabile che indica la distanza euclidea tra ciascun caso e il relativo centro di classificazione.

Cluster K-medie: Opzioni

Figura 34-5
Finestra di dialogo Cluster K-medie: Opzioni



Statistiche. È possibile selezionare le seguenti statistiche: centri iniziali del cluster, tabella ANOVA e informazioni sui cluster per ciascun caso.

- **Centri iniziali del cluster.** Prima stima delle medie delle variabili per ciascun cluster. In mancanza di indicazioni particolari, viene selezionato dai dati un numero di casi ben distanziati uguale al numero dei cluster. I centri dei cluster

iniziali vengono usati per un primo ciclo di classificazione e poi vengono aggiornati.

- **Tabella ANOVA (Cluster k-medie: opzioni).** Produce una tabella di analisi della varianza con test F per ogni variabile. I test F sono descrittivi e il livello di significatività fornisce informazioni utili. La tabella non viene creata se tutti i casi vengono assegnati a un solo cluster.
- **Cluster per ogni caso.** Visualizza per ogni caso il cluster di appartenenza e la distanza euclidea dal centro del cluster utilizzato per classificare il caso. Visualizza inoltre la distanza euclidea fra i centri finali.

Valori mancanti. Le opzioni disponibili sono Escludi casi listwise o Escludi casi pairwise.

- **Esclusione listwise.** Consente di escludere i casi coni valori mancanti per le variabili di raggruppamento dall'analisi.
- **Esclusione pairwise.** Consente di assegnare i casi ai cluster in base alle distanze calcolate da tutte le variabili con valori non mancanti.

Opzioni aggiuntive del comando QUICK CLUSTER

La procedura Cluster K-medie usa la sintassi del comando `QUICK CLUSTER`. Il linguaggio a comandi SPSS permette anche di:

- Accettare i primi *k* casi come centri dei cluster iniziali per evitare di dover leggere i dati normalmente usati per stimarli.
- Specificare i centri iniziali dei cluster direttamente come parte della sintassi del comando.
- Specificare i nomi delle variabili salvate.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
35

Test non parametrici

La procedura Test non parametrici è costituita da diversi test che non richiedono ipotesi sulla forma della distribuzione sottostante:

Test Chi-quadrato. Consente di analizzare una variabile in categorie e di ottenere una statistica del chi-quadrato in base alle differenze tra frequenze osservate e attese.

Test binomiale. Consente di confrontare la frequenza osservata in ciascuna categoria di una variabile dicotomica con le frequenze attese dalla distribuzione binomiale.

Test delle successioni. Consente di verificare se l'ordine di occorrenza di due valori di una variabile è casuale.

Test di Kolmogorov-Smirnov per un campione. Consente di confrontare la funzione di distribuzione cumulativa osservata per una variabile con la distribuzione teorica specificata, che può essere normale, uniforme esponenziale o di Poisson.

Test per due campioni indipendenti. Consente di confrontare due gruppi di casi in base a una sola variabile. Sono disponibili il test *U* di Mann-Whitney, il test di Kolmogorov-Smirnov per due campioni, il test delle reazioni estreme di Moses e il test delle successioni di Wald-Wolfowitz.

Test per due campioni dipendenti. Consente di confrontare le distribuzioni di due variabili. Sono disponibili il test di Wilcoxon, il test del segno e il test di McNemar.

Test per diversi campioni indipendenti. Consente di confrontare due o più gruppi di casi in base alla stessa variabile. Sono disponibili il test di Kruskal-Wallis, il test della mediana e il test di Jonckheere-Terpstra.

Test per diversi campioni dipendenti. Consente di confrontare le distribuzioni di due o più variabili. Sono disponibili il test di Friedman, il test *W* di Kendall e il test *Q* di Cochran.

Per tutti i test precedentemente citati sono disponibili quartili e media, deviazione standard, valore minimo e massimo e numero di casi non mancanti.

Test Chi-quadrato

La procedura Test chi-quadrato permette di analizzare una variabile in categorie e di calcolare una statistica chi-quadrato. Questo test sulla bontà di adattamento permette di confrontare le frequenze osservate e attese in ciascuna categoria per verificare se tutte le categorie includono la stessa proporzione di valori o se includono una proporzione di valori specificati dall'utente.

Esempio. Il test chi-quadrato può essere utilizzato per determinare se in un sacchetto di gelatine di frutta è presente la stessa proporzione di blu, marrone, arancio, rosso e giallo. È inoltre possibile determinare se il sacchetto di gelatine contiene il 5% di blu, il 30% di marrone, il 10% di verde, il 20% di arancio, il 15% di rosso e il 15% di giallo.

Statistiche. Media, deviazione standard, valore minimo e massimo e quartili. Il numero e la percentuale di casi mancanti e non mancanti, il numero di casi osservati e attesi per ciascuna categoria, i residui e la statistica chi-quadrato.

Dati. Utilizzare variabili categoriali numeriche ordinate o non ordinate (livelli di misurazione ordinali o nominali). Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale. Le frequenze attese per ciascuna categoria devono essere come minimo pari a 1. Non più del 20% delle categorie possono avere frequenze attese inferiori a 5.

Figura 35-1
Output del test chi-quadrato

Colore

	Numerosità osservata	Numerosità attesa	Residuo
Blu	6	18.8	-12.8
Marrone	33	18.8	14.2
Verde	9	18.8	-9.8
Giallo	17	18.8	-1.8
Arancione	22	18.8	3.2
Rosso	26	18.8	7.2
Totale	113		

Test

	Colore
Chi-quadrato[1]	27.973
df	5
Sig. Asint.	.000

1. Per 0 celle (.0%) erano previste frequenze minori di 5. Il valore minimo previsto per la frequenza in una cella è 18.8.

Colore

	Numerosità osservata	Numerosità attesa	Residuo
Blu	6	5.9	.1
Marrone	33	35.7	-2.7
Verde	9	11.9	-2.9
Giallo	17	23.8	-6.8
Arancione	22	17.8	4.2
Rosso	26	17.8	8.2
Totale	113		

Test

	Colore
Chi-quadrato[1]	7.544
df	5
Sig. Asint.	.183

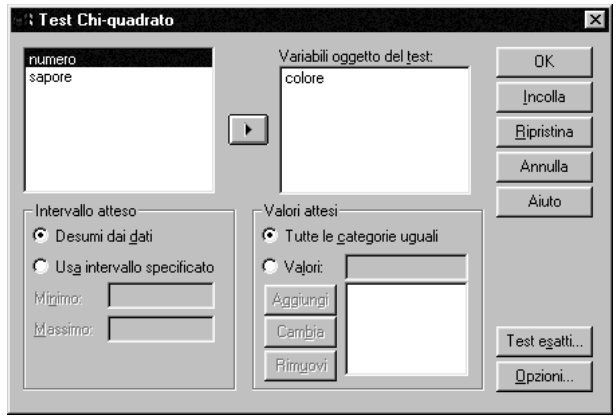
1. Per 0 celle (.0%) erano previste frequenze minori di 5. Il valore minimo previsto per la frequenza in una cella è 5.9.

Per ottenere un test chi-quadrato

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    Chi-quadrato...

Figura 35-2
Finestra di dialogo Test chi-quadrato



- Selezionare una o più variabili per il test. Ogni variabile produce un test distinto.
- È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Intervallo e valori attesi del test chi-quadrato

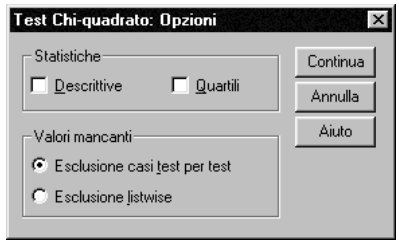
Intervallo atteso. Per impostazione predefinita, ogni singolo valore della variabile è definito come categoria. Per definire categorie all'interno di un intervallo specifico, selezionare Usaintervallo specificato e inserire valori interi per il limite inferiore e superiore. Le categorie verranno definite come valori inclusi nell'intervallo specificato. I casi con valori al di fuori del minimo e del massimo specificato saranno esclusi dal test. Se, ad esempio, si specifica 1 per il limite inferiore e 4 per il limite superiore, per il test chi-quadrato verranno utilizzati solo i valori interi compresi tra 1 e 4.

Valori attesi. Per impostazione predefinita, tutte le categorie hanno valori attesi uguali. Per le categorie sono previste proporzioni attese definite dall'utente. Selezionare Valori, specificare un valore maggiore di 0 per ogni categoria della variabile del test e quindi fare clic su Aggiungi. I valori vengono elencati in ordine di inserimento, L'ordine dei valori è importante in quanto corrisponde all'ordine crescente dei valori delle categorie della variabile oggetto del test. Il primo valore della lista corrisponde al valore di gruppo minore della variabile, mentre l'ultimo corrisponde al valore

maggiore. Gli elementi della lista dei valori vengono sommati e quindi ciascun valore viene diviso per la somma risultante per calcolare la proporzione di casi attesi nella categoria corrispondente. Ad esempio, una lista valori formata da 3, 4, 5, 4 specifica le proporzioni attese 3/16, 4/16, 5/16 e 4/16.

Test chi-quadrato: Opzioni

Figura 35-3
Finestra di dialogo Test chi-quadrato: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (test chi-quadrato)

Il linguaggio a comandi SPSS permette anche di:

- Specificare valori minimi e massimi diversi o frequenze attese diverse per variabili diverse (con il sottocomando `CHISQUARE`).
- Eseguire il test confrontando la stessa variabile con diverse frequenze attese o utilizzando diversi intervalli (con il sottocomando `EXPECTED`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test binomiale

Grazie alla procedura del test binomiale è possibile confrontare le frequenze osservate delle due categorie di una variabile dicotomica con le frequenze previste in presenza di una distribuzione binomiale con il parametro di probabilità specificato. Per impostazione predefinita, il parametro di probabilità per entrambi i gruppi è 0,5. Per modificare le probabilità, è possibile inserire una proporzione di prova per il primo gruppo. La probabilità per il secondo gruppo sarà uguale a 1 meno la probabilità specificata per il primo gruppo.

Esempio. Quando si lancia in aria una moneta, la probabilità che esca testa è pari a 1/2. In base a questa ipotesi, la moneta viene lanciata in aria 40 volte e i risultati (testa o croce) vengono registrati. Dal test binomiale può risultare che per i 3/4 dei lanci della moneta è uscita testa e che il livello di significatività è molto basso (0,0027). Questi risultati indicano che la probabilità che venga testa molto spesso non è pari a 1/2; pertanto, la stima sul comportamento della moneta probabilmente risulta distorta.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Dati. Le variabili incluse nel test devono essere numeriche e dicotomiche. Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma. Una **variabile dicotomica** è una variabile che prevede solo due possibili valori: *Sì* o *No*, *Vero* o *Falso*, 0 o 1 e così via. Se le variabili sono dicotomiche, è necessario specificare un punto di divisione. Utilizzando il punto di divisione è possibile assegnare a un gruppo i casi con valori superiori al punto di divisione e i rimanenti casi a un altro gruppo.

Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Si presume che i dati rappresentino un campione casuale.

Figura 35-4
Output del test binomiale

Test binomiale

		Categoria	Numerosità	Proporzione osservata.	Probabilità test.	Sig. Asint. a 2 code
Moneta	Gruppo 1	Testa	30	.75	.50	.003[1]
	Gruppo 2	Croce	10	.25		
	Totale		40	1.00		

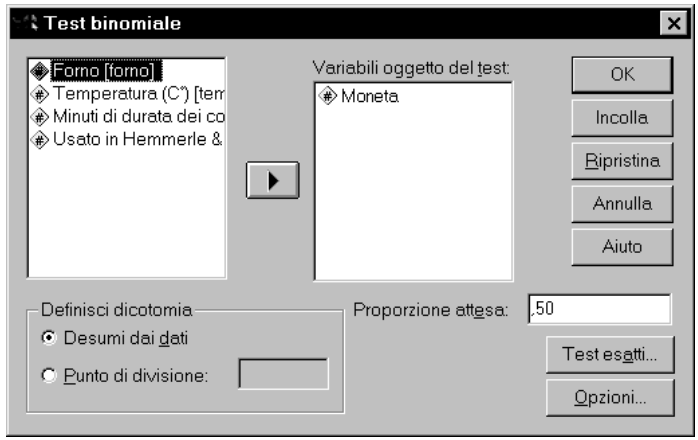
1. Approssimazione basata su Z.

Per ottenere un test binomiale

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    Binomiale...

Figura 35-5
Finestra di dialogo Test binomiale



▶ Selezionare una o più variabili numeriche oggetto del test.

▶ È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test binomiale: Opzioni

Figura 35-6
Finestra di dialogo Test binomiale: Opzioni

Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile da verificare verranno esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TEST (test binomiale)

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Selezionare gruppi specifici (ed escluderne altri) quando una variabile ha più di due categorie (con il sottocomando `BINOMIAL`).
- Specificare diversi punti di divisione o probabilità per variabili diverse (con il sottocomando `BINOMIAL`).
- Eseguire test confrontando la stessa variabile con diversi punti di divisione o probabilità (con il sottocomando `EXPECTED`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test delle successioni

Il test delle successioni verifica se l'ordine delle occorrenze di due valori di una variabile è casuale. Una successione è una sequenza di osservazioni simili. Un campione con troppe o troppo poche successioni indica che il campione non è casuale.

Esempio. Si supponga che a venti persone venga chiesto se comprerebbero un determinato prodotto. La casualità prevista per il campione viene messa fortemente in dubbio se tutte le venti persone sono dello stesso sesso. È possibile utilizzare il test delle successioni per determinare se il campione è stato definito in modo casuale.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Dati. Le variabili devono essere numeriche. Per convertire le variabili stringa in variabili numeriche, utilizzare il comando Ricodifica automatica del menu Trasforma.

Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Utilizzare campioni da distribuzioni di probabilità continue.

Figura 35-7
Output del test delle successioni

Test delle successioni

	Sesso
Valore test[1]	1,00
Casi < Valore test	7
Casi >= Valore test	13
Casi totali	20
Numero di successioni	15
Z	2,234
Significatività asintotica (2-code)	,025

1. Mediana

Per ottenere un test delle successioni

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    Successioni...

Figura 35-8
Finestra di dialogo Test delle successioni

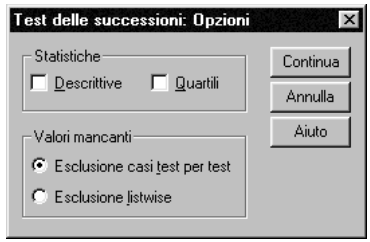
- Selezionare una o più variabili numeriche oggetto del test.
- È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test delle successioni: Punto di divisione

Punto di divisione. Specifica un punto di divisione per dicotomizzare le variabili scelte dall'utente. È possibile utilizzare la media osservata, la mediana o la moda oppure un valore specificato come punto di divisione. I casi con valori minori del punto di divisione sono assegnati a un gruppo e i casi con valori uguali o maggiori del punto di divisione sono assegnati a un altro gruppo. Viene eseguito un test per ogni punto di divisione selezionato.

Test delle successioni: Opzioni

Figura 35-9
Finestra di dialogo Test delle successioni: Opzioni



Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Test delle successioni)

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Specificare diversi punti di divisione per diverse variabili (con il sottocomando `RUNS`).
- Verificare la stessa variabile rispetto a diversi punti di divisione personalizzati (con il sottocomando `RUNS`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test di Kolmogorov-Smirnov per un campione

La procedura Test di Kolmogorov-Smirnov per un campione consente di confrontare la funzione di distribuzione cumulata osservata per una variabile con la distribuzione teorica specificata, che può essere normale, uniforme o di Poisson. La *Z* di Kolmogorov-Smirnov viene calcolata in base alla differenza maggiore (in valore assoluto) tra la funzione di distribuzione cumulata osservata e teorica. Questo test sulla bontà di adattamento permette di verificare se le osservazioni possono provenire dalla distribuzione specificata.

Esempio. Molti test parametrici richiedono variabili distribuite in modo normale. Il test di Kolmogorov-Smirnov per un campione può essere utilizzato per verificare che una variabile, ad esempio *reddito*, sia distribuita in modo normale.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili.

Dati. Utilizzare variabili quantitative (misurazione a livello di intervallo o di rapporto).

Assunzioni. Il test di Kolmogorov-Smirnov presume che i parametri della distribuzione del test vengano specificati anticipatamente. Questa procedura consente di valutare i parametri del campione. La media e la deviazione standard del campione sono i parametri della distribuzione normale, i valori minimo e massimo del campione definiscono l'intervallo di distribuzione uniforme e la media del campione è il parametro per la distribuzione Poisson e per la distribuzione esponenziale. La capacità del test di rilevare gli scostamenti dalla distribuzione ipotizzata può essere seriamente compromessa. Per effettuare test su una distribuzione normale con parametri stimati, è generalmente consigliabile usare il test K-S di Lilliefors corretto (selezionabile dalla procedura Esplora).

Figura 35-10
Output del test di Kolmogorov-Smirnov per un campione

Test di Kolmogorov-Smirnov per un campione

		Reddito
Numerosità		20
Parametri normali[1,2]	Media	56250.00
	Deviazione standard	45146.40
Differenze più estreme	Assoluto	.170
	Positivo	.170
	Negativo	-.164
Z di Kolmogorov-Smirnov		.760
Sig. Asint. a 2 code		.611

1. La distribuzione del test è Normale.
2. Calcolato dai dati.

Per ottenere un test di Kolmogorov-Smirnov per un campione

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    K-S per 1 campione...

Figura 35-11
Finestra di dialogo Test di Kolmogorov-Smirnov per un campione

- Selezionare una o più variabili numeriche oggetto del test. Ogni variabile produce un test distinto.
- È possibile fare clic su Opzioni per ottenere statistiche descrittive, quartili e controllo delle modalità di elaborazione dei dati mancanti.

Test di Kolmogorov-Smirnov per un campione: Opzioni

Figura 35-12
Finestra di dialogo K-S per un campione

Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Test di Kolmogorov-Smirnov per un campione)

Il linguaggio a comandi di SPSS consente anche di specificare i parametri della distribuzione del test (con il sottocomando `K-S`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test per due campioni indipendenti

La procedura del test per due campioni indipendenti consente di confrontare due gruppi di casi in base a una sola variabile.

Esempio. Sono stati creati nuovi apparecchi odontoiatrici che presentano numerosi vantaggi in termini di comodità, estetica ed efficacia ai fini dell'allineamento dei denti. Per determinare se i nuovi e i vecchi apparecchi devono essere portati per lo stesso periodo, sono stati scelti casualmente 10 bambini con il vecchio apparecchio e 10 bambini con il nuovo apparecchio. Dal test *U* di Mann-Whitney è possibile riscontrare che in media il nuovo apparecchio deve essere portato per un periodo di tempo inferiore.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: test *U* di Mann-Whitney, reazioni estreme di Moses, test *Z* di Kolmogorov-Smirnov, test delle successioni di Wald-Wolfowitz.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Assunzioni. Utilizzare campioni casuali indipendenti. Il test *U* di Mann-Whitney richiede l'utilizzo di due campioni di forma analoga.

Figura 35-13
Output per due campioni indipendenti

Ranghi

	Apparecchio	Numerosità	Rango medio	Somma dei ranghi
Giorni di utilizzo	Vecchio tipo	10	14.10	141.00
	Nuovo tipo	10	6.90	69.00
	Totale	20		

Test [2]

	Giorni di utilizzo
U di Mann-Whitney	14.000
W di Wilcoxon	69.000
Z	-2.721
Sig. Asint. a 2 code	.007
Significatività esatta [2*(Significatività a 1 coda)]	.005[1]

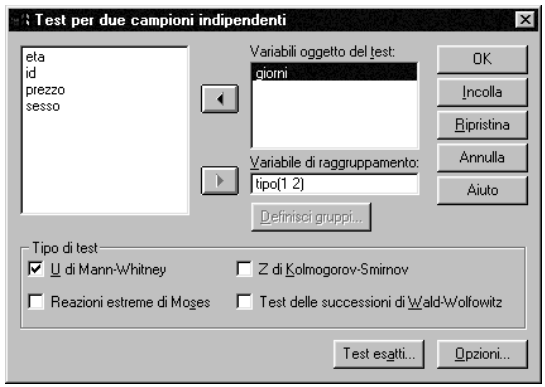
1. Non corretto per valori pari merito.
2. Variabile di raggruppamento: Apparecchio

Per ottenere un test per due campioni indipendenti

▶ Dai menu, scegliere:

Analizza
Test non parametrici
2 campioni indipendenti.

Figura 35-14
Finestra di dialogo Test per due campioni indipendenti



▶ Selezionare una o più variabili numeriche.

▶ Selezionare una variabile di raggruppamento e fare clic su Definisci gruppi per suddividere il file in due gruppi o campioni.

Tipi di test per due campioni indipendenti

Tipo di test. Sono disponibili quattro test che consentono di verificare se due campioni (gruppi) indipendenti provengono dalla stessa popolazione.

Il **test U di Mann-Whitney** è il test per due campioni indipendenti più diffuso. Equivale al test di Wilcoxon e al test di Kruskal-Wallis per due gruppi. Il test di Mann-Whitney permette di verificare l'equivalenza della posizione delle due popolazioni campione. Le osservazioni di entrambi i gruppi vengono combinate e ordinate per ranghi, assegnando la media dei ranghi ai valori a pari merito. Il numero di valori a pari merito deve essere inferiore al numero totale di osservazioni. Se la posizione delle popolazioni risulta identica, è necessario distribuire casualmente i ranghi tra i due campioni. Il test calcola il numero di volte in cui il punteggio del gruppo 1 è inferiore a quello del gruppo 2 e il numero di volte che un punteggio del gruppo 2 è inferiore al punteggio del gruppo 1. Il dato statistico che si ottiene con il test *U* di Mann-Whitney è inferiore a questi due numeri. Verrà visualizzata anche la statistica risultante dal test *W* di Wilcoxon, corrispondente alla somma dei ranghi del campione più piccolo. Se per entrambi i campioni è stato eseguito lo stesso numero di osservazioni, *W* è la somma dei ranghi del gruppo elencato per primo nella finestra di dialogo Due campioni indipendenti: Definisci gruppi.

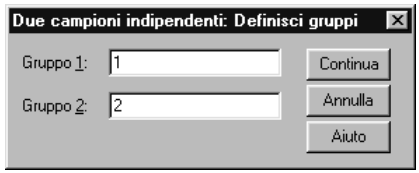
Il **test Z di Kolmogorov-Smirnov** e il **test delle successioni di Wald-Wolfowitz** sono test di carattere più generale che consentono di individuare le differenze tra le distribuzioni in termini di forma e posizione. Il test di Kolmogorov-Smirnov si basa sulla massima differenza in valore assoluto tra le funzioni di distribuzione cumulative osservate per entrambi i campioni. Quando tale differenza è significativa, le due distribuzioni vengono considerate diverse. Il test delle successioni di Wald-Wolfowitz consente di combinare e ordinare in ranghi le osservazioni di entrambi i gruppi. Se i due campioni provengono dalla stessa popolazione, è necessario distribuire casualmente i gruppi all'interno della classifica.

Il **test delle reazioni estreme di Moses** si basa sull'ipotesi che la variabile sperimentale influenzi alcuni soggetti in una direzione e altri nella direzione opposta. Consente di verificare le risposte estreme confrontandole con un gruppo di controllo. Questo test è incentrato sull'estensione del gruppo di controllo e definisce la misura in cui i valori estremi del gruppo sperimentale influenzano l'estensione in caso di combinazione con il gruppo di controllo. Il gruppo di controllo viene definito dal valore del gruppo 1 nella finestra di dialogo Due campioni indipendenti: Definisci gruppi. Le osservazioni eseguite su entrambi i gruppi vengono combinate e classificate per ranghi. L'estensione del gruppo di controllo viene calcolata come la

differenza tra i ranghi dei valori massimo e minimo del gruppo di controllo più 1. Poiché a causa di valori casuali anomali è probabile che l'intervallo dell'estensione risulti distorto, da ciascun estremo viene eliminato il 5% dei casi di controllo.

Test per due campioni indipendenti: Definisci gruppi

Figura 35-15
Finestra di dialogo Due campioni indipendenti: Definisci gruppi



Per suddividere il file in due gruppi o campioni, inserire un valore intero per il gruppo 1 e un altro per il gruppo 2. I casi con altri valori verranno esclusi dall'analisi.

Test per due campioni indipendenti: Opzioni

Figura 35-16
Finestra di dialogo Due campioni indipendenti: Opzioni

Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (Due campioni indipendenti)

Il linguaggio a comandi SPSS permette anche di specificare il numero di casi da eliminare dal test di Moses (con il sottocomando `MOSES`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test per due campioni dipendenti

La procedura del test per due campioni dipendenti consente di confrontare la distribuzione di due variabili.

Esempio. In generale, le famiglie ricevono l'intero prezzo di offerta per la vendita della propria casa? Applicando il test di Wilcoxon ai dati relativi a 10 case, si riscontrerà che sette famiglie ricevono una somma inferiore al prezzo di offerta, una famiglia riceve una somma superiore, mentre due sole famiglie lo ricevono interamente.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: di Wilcoxon, del segno e di McNemar.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Assunzioni. Anche se per le due variabili non si ipotizza una particolare distribuzione, si presume che la distribuzione delle differenze a coppie sia simmetrica.

Figura 35-17
Output di due campioni dipendenti

Ranghi

		Numerosità	Rango medio	Somma dei ranghi
Prezzo di vendita - Prezzo di listino	Ranghi negativi	7[1]	4.93	34.50
	Ranghi positivi	1[2]	1.50	1.50
	Valori pari merito	2[3]		
	Totale	10		

1. Prezzo di vendita < Prezzo di listino
2. Prezzo di vendita > Prezzo di listino
3. Prezzo di listino = Prezzo di vendita

Test[2]

	Prezzo di vendita - Prezzo di listino
Z	-2.313[1]
Sig. Asint. a 2 code	.021

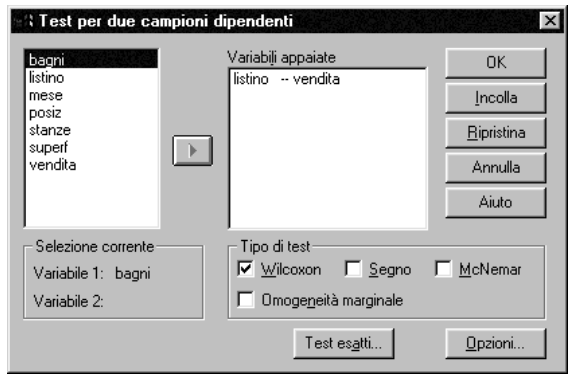
1. Basato su ranghi positivi.
2. Test di Wilcoxon

Per ottenere test per due campioni dipendenti

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    2 campioni dipendenti.

Figura 35-18
Finestra di dialogo Test per due campioni dipendenti



- Selezionare una o più coppie di variabili, come segue:
  - Fare clic su ciascuna variabile. La prima variabile comparirà nel gruppo Selezione corrente come *Variabile 1*, mentre la seconda comparirà come *Variabile 2*.
  - Fare clic sul pulsante freccia per spostare la coppia nell'elenco Variabili appaiate. È possibile selezionare più coppie di variabili. Per rimuovere una coppia di variabili dall'analisi, selezionarla nell'elenco Variabili appaiate e quindi fare clic sul pulsante freccia.

Tipi di test per due campioni dipendenti

I test descritti in questa sezione permettono di confrontare le distribuzioni di due variabili correlate. Il test più appropriato varia a seconda dei tipi di dati.

Se i dati sono continui, utilizzare il test del segno o di Wilcoxon. Il **test del segno** permette di calcolare le differenze tra le due variabili per tutti i casi e di classificarle come positive, negative o a pari merito. Se le due variabili sono distribuite in modo analogo, il numero di differenze positive e negative non differirà in misura significativa. Il **test di Wilcoxon** prende in considerazione le informazioni relative al segno e all'entità delle differenze tra le coppie. Poiché il test di Wilcoxon include un maggior numero di informazioni relative ai dati, risulta più valido del test del segno.

Se i dati sono binari, utilizzare il **test di McNemar**. Questo test viene in genere utilizzato in presenza di misure ripetute, ovvero quando la risposta del soggetto viene richiesta due volte: prima e dopo il verificarsi di un determinato evento. Il test di McNemar consente di determinare se il tasso di risposta iniziale (prima dell'evento) equivale al tasso di risposta finale (dopo l'evento). Questo test risulta particolarmente utile per individuare le variazioni della risposta in disegni sperimentali del tipo 'prima e dopo'.

Se i dati sono categoriali, utilizzare il **test di omogeneità marginale**. Estensione del test di McNemar dalla risposta binaria a quella multinomiale. Consente di verificare le variazioni della risposta utilizzando la distribuzione del chi-quadrato e risulta utile in disegni sperimentali del tipo 'prima e dopo'. Il test di omogeneità marginale è disponibile solo se è stato installato il modulo Exact Tests.

Test per due campioni dipendenti: Opzioni

Figura 35-19
Finestra di dialogo Due campioni dipendenti: Opzioni

Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (due campioni dipendenti)

Il linguaggio a comandi SPSS permette anche di verificare una variabile con ciascuna variabile dell'elenco.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test per diversi campioni indipendenti

La procedura per i test per diversi campioni indipendenti consente di confrontare due o più gruppi di casi in base a una variabile.

Esempio. Le lampadine da 100 watt di tre diversi produttori si differenziano in relazione al tempo medio di bruciatura del filamento? Grazie all'ANOVA univariata di Kruskal-Wallis è possibile verificare che la durata media delle tre lampadine è effettivamente diversa.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: *H* di Kruskal-Wallis, mediana.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Assunzioni. Utilizzare campioni casuali indipendenti. IL test *H* di Kruskal-Wallis richiede che i campioni sottoposti a test siano simili per forma.

Figura 35-20
Output del test per diversi campioni indipendenti

Ranghi

			N	Rango Medio
Ore	Marca	Marca A	10	15,20
		Marca B	10	25,50
		Marca C	10	5,80
		Totale	30	

Test [1,2]

	Ore di luce
Chi-quadrato	25.061
df	2
Sig. Asint.	.000

1. Test di Kruskal Wallis
2. Variabile di raggruppamento: Marca

Per ottenere test per diversi campioni indipendenti

▶ Dai menu, scegliere:

Analizza
  Test non parametrici
    K campioni indipendenti...

Figura 35-21
Finestra di dialogo Test per diversi campioni indipendenti

- Selezionare una o più variabili numeriche.
- Selezionare una variabile di raggruppamento e fare clic su Definisci intervallo per specificare i valori interi minimo e massimo per la variabile di raggruppamento.

Test per diversi campioni indipendenti: tipi di test

Sono disponibili tre test per stabilire se diversi campioni indipendenti sono stati estratti dalla stessa popolazione. Il test *H* di Kruskal-Wallis, il test della mediana e il test di Jonckheere-Terpstra consentono di verificare se i diversi campioni indipendenti sono stati estratti dalla stessa popolazione.

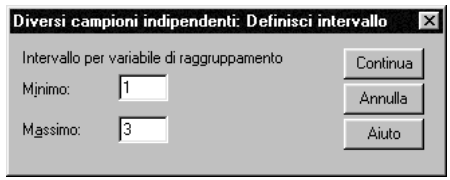
Il test **H di Kruskal-Wallis**, un'estensione del test *U* di Mann-Whitney, è la versione non parametrica dell'analisi univariata della varianza e consente di rilevare le differenze nella posizione di distribuzione. Il **test della mediana**, che è più generale ma non altrettanto potente, consente di rilevare le differenze distribuzionali nella posizione e nella forma. Il test *H* di Kruskal-Wallis e il test della mediana presumono che non esistano ordinamenti *a priori* delle *k* popolazioni da cui sono estratti i campioni.

Quando *esiste* un naturale ordinamento *a priori* (crescente o decrescente) delle *k* popolazioni, il **test di Jonckheere-Terpstra** è più potente. Ad esempio, le *k* popolazioni possono rappresentare *k* temperature crescenti. L'ipotesi che diverse temperature producano la stessa distribuzione della risposta è verificata rispetto all'ipotesi alternativa in base a cui al salire della temperatura, cresce il valore della risposta. Qui l'ipotesi alternativa è ordinata e quindi il test di Jonckheere-Terpstra è il

più appropriato da utilizzare. Il test di Jonckheere-Terpstra è disponibile solo se è installato il modulo Exact Tests.

Test per diversi campioni indipendenti: Definisci intervallo

Figura 35-22
Finestra di dialogo Diversi campioni indipendenti: Definisci intervallo



Per definire l'intervallo, immettere valori interni per il minimo e il massimo che corrispondono alle categorie minore e maggiore della variabile di raggruppamento. Sono esclusi i casi con valori al di fuori dei limiti. Se, ad esempio, si specifica un limite inferiore di 1 e un limite superiore di 3, verranno utilizzati solo i valori interi compresi tra 1 e 3. Il valore minimo deve essere inferiore al valore massimo ed entrambi i valori devono essere specificati.

Test per diversi campioni indipendenti: Opzioni

Figura 35-23
Finestra di dialogo Diversi campioni indipendenti: Opzioni

Statistiche. È possibile scegliere uno o entrambe le statistiche riassuntive.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Valori mancanti. Consente di controllare la modalità di elaborazione dei valori mancanti.

- **Esclusione casi test per test.** Quando vengono specificati più test, in ciascuno verranno valutati separatamente i valori mancanti.
- **Esclusione listwise.** I casi con valori mancanti per qualsiasi variabile sono esclusi da tutte le analisi.

Funzioni aggiuntive del comando NPAR TESTS (K campioni indipendenti)

Il linguaggio a comandi SPSS permette anche di specificare un valore diverso dalla mediana osservata per il test della mediana (con il sottocomando `MEDIAN`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Test per diversi campioni dipendenti

Il test per diversi campioni dipendenti consente di confrontare le distribuzioni di due o più variabili.

Esempio. Il pubblico associa diversi livello di prestigio al ruolo di dottore, avvocato, ufficiale della polizia e insegnante? A dieci persone viene chiesto di ordinare queste quattro occupazioni in base al prestigio. Il test di Friedman indica che il pubblico associa effettivamente livelli di prestigio diversi a queste quattro professioni.

Statistiche. Media, deviazione standard, valore minimo, valore massimo, numero di casi non mancanti e quartili. Test: Friedman, *W* di Kendall e *Q* di Cochran.

Dati. Utilizzare variabili numeriche che possono essere ordinate.

Assunzioni. I test non parametrici non richiedono ipotesi relative alla forma della distribuzione sottostante. Utilizzare campioni casuali dipendenti.

Figura 35-24
Output del test per diversi campioni dipendenti

Ranghi

	Rango medio
Medico	1.50
Avvocato	2.50
Poliziotto	3.40
Insegnante	2.60

Test[1]

Numerosità	10
Chi-quadrato	10.920
df	3
Sig. Asint.	.012

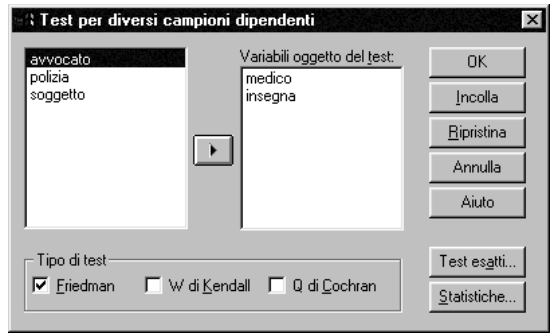
1. Test di Friedman

Per ottenere i test per diversi campioni dipendenti

- Dai menu, scegliere:

Analizza
  Test non parametrici
    K campioni dipendenti...

Figura 35-25
Finestra di dialogo Test per diversi campioni dipendenti



- Selezionare una o più variabili oggetto del test numeriche.

Test per diversi campioni dipendenti: tipi di test

Sono disponibili tre test per confrontare le distribuzioni di diverse variabili correlate.

Il **testdi Friedman** è l'equivalente non parametrico di un disegno di misure ripetute per un campione o ANOVA a due vie con una osservazione per cella. Friedman verifica l'ipotesi nulla secondo cui *k* variabili correlate provengono dalla stessa popolazione. Per ogni caso, alle *k* variabili viene assegnato un rango da 1 a *k*. Le statistiche del test sono basate su questi ranghi.

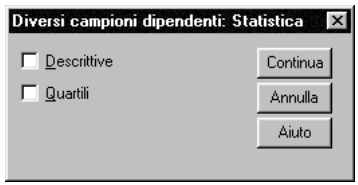
W di Kendall è una normalizzazione delle statistiche di Friedman. È possibile interpretare il *W* di Kendall come il coefficiente di concordanza, che rappresenta la misura dell'accordo tra stimatori. Ogni caso è uno stimatore e ogni variabile è un

elemento o individuo da stimare. Per ogni variabile viene calcolata la somma dei ranghi. *W* di Kendall varia tra 0 (nessun accordo) e 1 (accordo completo).

Q di Cochran è identico al test di Friedman ma è applicabile quando tutte le risposte sono binarie. Questo test è un'estensione del test di McNemar alla situazione di *k*-campioni. I test *Q* di Cochran verificano l'ipotesi secondo cui diverse variabili dicotomiche hanno la stessa media. Le variabili sono misurate sullo stesso individuo o su individui collegati fra loro.

Test per diversi campioni dipendenti: Statistica

Figura 35-26
Finestra di dialogo Diversi campioni dipendenti: Statistica



È possibile scegliere le statistiche.

- **Descrittive.** Consente di visualizzare la media, la deviazione standard, il valore minimo e massimo e il numero di casi non mancanti.
- **Quartili.** Consente di visualizzare i valori corrispondenti al 25°, 50° e 75° percentile.

Funzioni aggiuntive del comando NPAR TESTS (K campioni dipendenti)

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
36

Analisi a risposta multipla

Sono disponibili due procedure per l'analisi di insiemi a dicotomie e a categorie multiple. La procedura Risposte multiple: Frequenze consente di visualizzare le tabelle di frequenza. La procedura Risposte multiple: Tavole di contingenza consente di visualizzare tavole di contingenza a due o a tre dimensioni. Prima di utilizzare una delle procedure descritte, è necessario definire gli insiemi a risposta multipla.

Esempio. In questo esempio viene illustrato l'utilizzo degli elementi a risposta multipla in un'indagine di mercato. I dati sono fittizi e non devono essere interpretati come reali. È possibile condurre un'indagine tra i passeggeri di una linea aerea in volo su una particolare rotta per ottenere una valutazione della concorrenza. In questo esempio la compagnia American Airlines conduce un'indagine volta a rilevare se i propri passeggeri viaggiano con altre linee aeree sulla rotta Chicago-New York e a determinare l'importanza relativa dei fattori di programmazione e di servizio ai fini della scelta della linea aerea. Al momento dell'imbarco, l'assistente di volo consegna a ciascun passeggero un breve questionario. La prima domanda è la seguente. Tra le seguenti compagnie aeree, contrassegnare quelle utilizzate su questa stessa rotta almeno una volta negli ultimi sei mesi: American, United, TWA, USAir, Altro. Si tratta di una domanda a risposta multipla in quanto il passeggero può indicare più risposte. La domanda, tuttavia, non può essere codificata direttamente in quanto una variabile può contenere un solo valore per ciascun caso. È necessario utilizzare più variabili per associare le risposte a ciascuna domanda. Per eseguire questa operazione è possibile procedere in due modi. Il primo modo consiste nel definire una variabile per ciascuna delle scelte (ad esempio, American, United, TWA, USAir e Altro). Se il passeggero indica la linea United, alla variabile *united* verrà assegnato il codice 1 e in caso contrario il codice 0. Si tratta di un **metodo a dicotomie multiple** per l'assegnazione delle variabili. Il secondo metodo per la classificazione delle risposte è il **metodo a categorie multiple**, che consente di valutare il numero massimo di risposte possibili alla domanda e di impostare lo stesso numero di variabili, con i codici utilizzati per specificare la linea aerea utilizzata. Dall'esame di un campione

di questionari può risultare che negli ultimi sei mesi nessun utente ha viaggiato con più di tre diverse linee aeree su questa rotta. Può inoltre risultare che, a causa della deregulation delle linee aree, nella categoria Altro ne vengano citate altre 10. Utilizzando il metodo a risposta multipla, vengono definite tre variabili, ciascuna delle quali è codificata come 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta* e così via. Se un determinato passeggero indica American e TWA, alla prima variabile viene assegnato il codice 1, alla seconda il codice 3 e alla terza un codice di valore mancante. Un altro passeggero può aver indicato American e specificato Delta. In questo caso, alla prima variabile viene assegnato il codice 1, alla seconda 5 e alla terza un codice di valore mancante. Se invece si utilizza il metodo a dicotomie multiple, si otterranno 14 variabili distinte. Sebbene ai fini di questa indagine sia possibile utilizzare entrambi i metodi, la scelta del metodo dipende dalla distribuzione delle risposte.

Risposte multiple: Definisci insiemi

La procedura di definizione degli insiemi di variabili a risposta multipla consente di raggruppare le variabili elementari in insiemi a dicotomie o a categorie multiple, per i quali è possibile ottenere tabelle di frequenza e tavole di contingenza. È possibile definire fino a 20 insiemi a risposta multipla. A ogni insieme è necessario assegnare un nome univoco. Per rimuovere un insieme, evidenziarlo nell'elenco dei gruppi a risposta multipla e quindi scegliere Rimuovi. Per modificare un insieme, evidenziarlo nell'elenco, modificare le caratteristiche di definizione desiderate e quindi scegliere Cambia.

È possibile codificare le variabili elementari come dicotomie o categorie. Per l'utilizzo delle variabili dicotomiche, selezionare Dicotomie per creare un insieme a dicotomie multiple. Specificare un valore intero nella casella Valore conteggiato. Ciascuna variabile contenente almeno un'occorrenza del valore conteggiato diventa una categoria dell'insieme a dicotomie multiple. Selezionare Categorie per creare un insieme a categorie multiple con lo stesso intervallo di valori delle variabili che lo compongono. Specificare valori interi come valori minimo e massimo dell'intervallo di categorie dell'insieme a categorie multiple. Verrà calcolato il totale di ogni singolo valore intero nell'intervallo per tutte le variabili. Le categorie vuote non verranno incluse nella tabella.

A ogni insieme a risposta multipla deve essere assegnato un nome univoco composto al massimo da sette caratteri. Al nome assegnato verrà aggiunto automaticamente il prefisso $ (segno di dollaro). Non è possibile utilizzare i seguenti nomi riservati: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* e *width*. Il nome

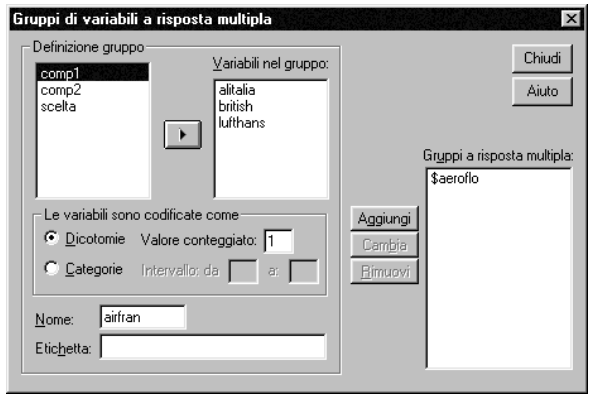
dell'insieme a risposta multipla esiste solo ai fini dell'utilizzo in procedure a risposta multipla. Non è possibile fare riferimento ai nomi di insiemi a risposta multipla in altre procedure. È inoltre possibile inserire un'etichetta di variabile descrittiva per l'insieme a risposta multipla. L'etichetta può essere costituita al massimo da 40 caratteri.

Per definire gli insiemi a risposta multipla

- Dai menu, scegliere:

  Analizza
  Risposte multiple
  Definisci insiemi...

Figura 36-1
Finestra di dialogo Insiemi a risposta multipla



- Selezionare due o più variabili.

- Se le variabili sono codificate come dicotomie, indicare il valore che si desidera calcolare. Se le variabili sono codificate come categorie, definire l'intervallo delle categorie.

- Immettere un nome univoco per ciascun insieme a risposta multipla.

- Scegliere Aggiungi per aggiungere l'insieme a risposta multipla all'elenco di insiemi definiti.

Risposte multiple: Frequenze

La procedura Risposte multiple: Frequenze consente di ottenere tabelle di frequenza per gli insiemi a risposta multipla. È innanzitutto necessario definire uno o più insiemi a risposta multipla (vedere "Risposte multiple: Definisci insiemi").

Per gli insiemi a dicotomie multiple, i nomi delle categorie indicati nell'output vengono determinati in base alle etichette definite per le variabili elementari del gruppo. Se le etichette delle variabili non sono definite, i nomi delle variabili verranno utilizzati come etichette. Per gli insiemi a categorie multiple, le etichette di categoria vengono determinate in base alle etichette dei valori della prima variabile del gruppo. Se le categorie mancanti per la prima variabile sono presenti per altre variabili del gruppo, definire un'etichetta dei valori per le categorie mancanti.

Valori mancanti. I casi con valori mancanti vengono esclusi tabella per tabella. È inoltre possibile scegliere una delle seguenti opzioni o entrambe:

- **Esclusione listwise all'interno delle dicotomie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a dicotomie multiple. Questa opzione può essere applicata solo agli insiemi a risposta multipla definiti come insiemi dicotomici. Per impostazione predefinita, un caso viene considerato mancante per un insieme a dicotomie multiple se nessuna delle variabili che lo compongono contiene il valore conteggiato. I casi con valori mancanti solo per alcune variabili verranno inclusi nelle tabelle del gruppo se almeno una variabile contiene il valore conteggiato.
- **Esclusione listwise all'interno delle categorie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a categorie multiple. Questa opzione viene applicata solo a insiemi a risposta multipla definiti come insiemi di categorie. Per impostazione predefinita, un caso viene considerato mancante per un insieme a categorie multiple solo se nessuno dei componenti contiene valori validi all'interno dell'intervallo definito.

Esempio. Ogni variabile creata in base a una domanda di un questionario è una variabile elementare. Per analizzare un elemento di un insieme a risposta multipla, è necessario unire le variabili in uno dei due tipi di insiemi a risposta multipla: un insieme a dicotomie multiple oppure un insieme a categorie multiple. Se ad esempio in un'indagine sulle linee aeree dove viene richiesto con quale delle tre linee aeree indicate (American, United, TWA) si è viaggiato negli ultimi sei mesi sono state utilizzate variabili dicotomiche e si è definito un **insieme a dicotomie multiple**, ciascuna delle tre variabili dell'insieme diventerà una categoria della variabile di

gruppo. I conteggi e le percentuali relativi alle tre linee aeree verranno visualizzati in una sola tabella di frequenza. Se risulta che nessuna persona ha indicato più di due linee aeree, è possibile creare due variabili, ciascuna con tre codici, ovvero uno per ogni linea aerea. Se si definisce un **insieme a categorie multiple**, i valori verranno inseriti nella tabella aggiungendo gli stessi codici alle variabili elementari. L'insieme di valori risultante equivale agli insiemi di ciascuna variabile elementare. Trenta risposte per United rappresentano ad esempio la somma delle cinque per United per la linea aerea 1 e delle venticinque risposte per United per la linea aerea 2. I conteggi e le percentuali relativi alle tre linee aeree verranno visualizzati in una sola tabella di frequenza.

Statistiche. Tabelle di frequenza in cui vengono visualizzati i conteggi, le percentuali delle risposte, le percentuali dei casi, il numero di casi validi e il numero di casi mancanti.

Dati. Utilizzare gli insiemi a risposta multipla.

Assunzioni. I conteggi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione.

Procedure correlate. La procedura Risposte multiple: Definisci insiemi consente di definire insiemi a risposta multipla.

Figura 36-2
Output della procedura Risposte multiple: Frequenze

		Risposte		Percentuale di casi
		N	Percentuale	
Nome[a]	American	75	67,6%	92,6%
	TWA	6	5,4%	7,4%
	United	30	27,0%	37,0%
Totale		111	100,0%	137,0%

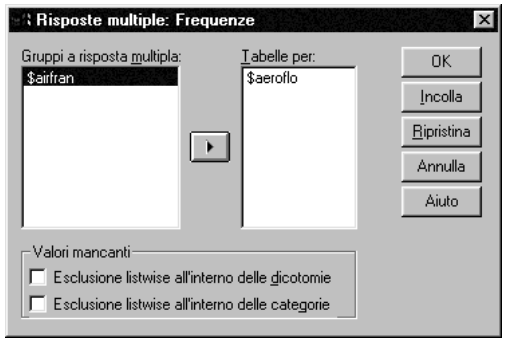
a. Gruppo a dicotomie incluso nella tabella al valore1.

Per ottenere le frequenze delle risposte multiple

▶ Dai menu, scegliere:

Analizza
Risposte multiple
Frequenze...

Figura 36-3
Finestra di dialogo Risposte multiple: Frequenze



- Selezionare uno o più insiemi a risposta multipla.

Risposte multiple: Tavole di contingenza

La procedura Risposte multiple: Tavole di contingenza consente di incrociare insiemi a risposta multipla definiti, variabili elementari o una combinazione di entrambi. È inoltre possibile ottenere le percentuali nelle celle in base a casi o risposte, modificare il trattamento dei valori mancanti oppure accoppiare le variabili di insiemi diversi. È innanzitutto necessario definire uno o più insiemi a risposta multipla (vedere "Per definire gli insiemi a risposta multipla").

Per gli insiemi a dicotomie multiple, i nomi delle categorie indicati nell'output vengono determinati in base alle etichette definite per le variabili elementari del gruppo. Se le etichette delle variabili non sono definite, i nomi delle variabili verranno utilizzati come etichette. Per gli insiemi a categorie multiple, le etichette di categoria vengono determinate in base alle etichette dei valori della prima variabile del gruppo. Se le categorie mancanti per la prima variabile sono presenti per altre variabili del gruppo, definire un'etichetta dei valori per le categorie mancanti. Le etichette delle categorie per le colonne verranno visualizzate su tre righe, composte al massimo da otto caratteri ciascuna. Per evitare di troncare le parole, è possibile invertire le righe e le colonne oppure ridefinire le etichette.

Esempio. In questa procedura è possibile incrociare gli insiemi a dicotomie e a categorie multiple con altre variabili. In un questionario riservato ai passeggeri delle linee aeree venivano richieste le seguenti informazioni: Contrassegnare tra le seguenti tutte le linee aeree di cui ci si è serviti almeno una volta negli ultimi sei

mesi (American, United, TWA). Cosa è più importante per la scelta di un volo, la programmazione o il servizio offerto? Scegliere una sola opzione. Dopo aver inserito i dati come dicotomie o categorie multiple e averli uniti in un insieme, è possibile incrociare le domande relative alle linee aeree con quelle relative al servizio o alla programmazione.

Statistiche. Tavole di contingenza con conteggi relativi a celle, righe, colonne e totale e percentuali relative a celle, righe, colonne e totale. Le percentuali nelle celle possono basarsi sui casi o sulle risposte.

Dati. Utilizzare insiemi a risposta multipla oppure variabili categoriali numeriche.

Assunzioni. I conteggi e le percentuali forniscono un'utile descrizione dei dati provenienti da qualsiasi distribuzione.

Procedure correlate. La procedura Risposte multiple: Definisci insiemi consente di definire insiemi a risposta multipla.

Figura 36-4
Output di tavole di contingenza a risposte multiple

			Linea aerea selezionata in base a		
			Programmazione	Servizio	Totale
Nome	American	Conteggio	41	34	75
	TWA	Conteggio	3	3	6
	United	Conteggio	27	3	30
Totale		Conteggio	44	37	81

Le percentuali e i totali si basano sui rispondenti.

a. Gruppo a dicotomie incluso nella tabella al valore1.

Per ottenere tavole di contingenza a risposte multiple

▶ Dai menu, scegliere:

Analizza
  Risposte multiple
    Tavole di contingenza...

Figura 36-5
Finestra di dialogo Risposte multiple: Tavole di contingenza

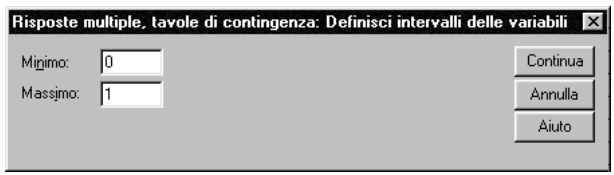
- Selezionare una o più variabili numeriche o insiemi a risposta multipla per ciascuna dimensione delle tavole di contingenza.

- Definire l'intervallo di ciascuna variabile elementare.

  È inoltre possibile ottenere una tavola di contingenza a due vie per ciascuna categoria di una variabile di controllo o di un insieme a risposta multipla. Selezionare uno o più elementi dall'elenco Strati.

Risposte multiple, tavole di contingenza: Definisci intervalli delle variabili

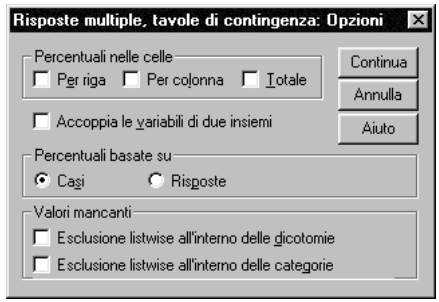
Figura 36-6
Finestra di dialogo Risposte multiple, tavole di contingenza: Definisci intervalli delle variabili



È necessario definire gli intervalli dei valori per tutte le variabili elementari delle tavole di contingenza. Specificare il valore di categoria minimo e massimo che si desidera inserire nelle tavole di contingenza. Le categorie che non rientrano nell'intervallo verranno escluse dall'analisi. Si assume che i valori inclusi nell'intervallo siano interi (i non interi verranno troncati).

Risposte multiple, tavole di contingenza: Opzioni

Figura 36-7
Finestra di dialogo Risposte multiple, tavole di contingenza: Opzioni



Percentuali nelle celle. I conteggi di cella vengono sempre visualizzati. È possibile impostare la visualizzazione di percentuali di riga, percentuali di colonna e percentuali per tabelle a due vie (totale).

Percentuali basate su. È possibile basare le percentuali nelle celle sui casi (o persone che rispondono). Questa opzione non è disponibile se si seleziona la corrispondenza delle variabili tra insiemi a categorie multiple. È inoltre possibile basare le percentuali nelle celle sulle risposte. Per gli insiemi a dicotomie multiple, il numero di risposte equivale al numero di valori conteggiati nei diversi casi. Per gli insiemi a categorie multiple, il numero di risposte equivale al numero di valori dell'intervallo definito.

Valori mancanti. È possibile scegliere una delle seguenti opzioni o entrambe:

- **Esclusione listwise all'interno delle dicotomie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a dicotomie multiple. Questa opzione può essere applicata solo agli insiemi a risposta multipla definiti come insiemi dicotomici. Per impostazione predefinita, un caso viene considerato mancante per un insieme a dicotomie multiple se nessuna delle variabili che lo compongono contiene il valore conteggiato. I casi con valori mancanti solo per alcune variabili verranno inclusi nelle tabelle del gruppo se almeno una variabile contiene il valore conteggiato.
- **Esclusione listwise all'interno delle categorie.** Consente di escludere i casi con valori mancanti per qualsiasi variabile dalla tabella dell'insieme a categorie multiple. Questa opzione viene applicata solo a insiemi a risposta multipla definiti come insiemi di categorie. Per impostazione predefinita, un caso viene

considerato mancante per un insieme a categorie multiple solo se nessuno dei componenti contiene valori validi all'interno dell'intervallo definito.

Per impostazione predefinita, quando si incrociano due insiemi a categorie multiple, ciascuna variabile del primo gruppo verrà incrociata con ciascuna variabile del secondo gruppo e quindi verranno sommati i conteggi relativi a ciascuna cella. Alcune risposte, pertanto, potranno comparire più volte nella stessa tabella. È possibile scegliere la seguente opzione:

Accoppia le variabili di due insiemi. Consente di associare la prima variabile del primo gruppo con la prima variabile del secondo gruppo e così via. Se viene selezionata questa opzione, le percentuali nelle celle saranno basate sulle risposte e non sulle persone che rispondono. Questa opzione non è disponibile per gli insiemi a dicotomie multiple né per le variabili elementari.

Funzioni aggiuntive del comando MULT RESPONSE

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Ottenere tavole di contingenza con un massimo di cinque dimensioni (con il sottocomando `BY`).
- Modificare le opzioni di formattazione dell'output, inclusa l'eliminazione delle etichette dei valori (con il sottocomando `FORMAT`).

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
37

Report

Risultati di report

Gli elenchi dei casi e le statistiche descrittive sono strumenti fondamentali per lo studio e la presentazione dei dati. Per creare elenchi dei casi è possibile utilizzare l'Editor dei dati o la procedura Riassumi, per produrre conteggi di frequenze e statistiche descrittive è possibile utilizzare la procedura Frequenze, mentre per creare statistiche per la sottopopolazione è possibile utilizzare la procedura Medie. Queste procedure utilizzano un formato progettato per rendere chiare le informazioni. Per visualizzare le informazioni in un formato diverso, è possibile impostare la presentazione dei dati mediante le opzioni per i report Riepiloghi per righe e Riepiloghi per colonne.

Report: Riepiloghi per righe

La procedura Report: Riepiloghi per righe consente di creare report in cui statistiche riassuntive diverse sono disposte in righe distinte. Sono inoltre disponibili gli elenchi dei casi, che possono includere o meno le statistiche riassuntive.

Esempio. Una ditta proprietaria di una catena di negozi al dettaglio registra le informazioni sui dipendenti, che includono informazioni sugli stipendi e le mansioni nonché sul negozio e il reparto in cui lavora ogni dipendente. È quindi possibile creare un report che includa le informazioni relative a ogni impiegato (elenco) suddivise per negozio e per reparto (variabili di separazione) e che includa statistiche riassuntive (ad esempio, lo stipendio medio) per ciascun negozio e reparto nonché per ciascun reparto di ogni negozio.

Colonne dati. Elenca le variabili da rappresentare nel report, per le quali si desidera creare elenchi dei casi o statistiche riassuntive e controlla il formato per la visualizzazione delle colonne di dati.

Colonne di separazione. Elenca le variabili di separazione facoltative che suddividono il report in più gruppi e consente di gestire le statistiche riassuntive e i formati per la visualizzazione delle colonne di separazione. Se sono presenti più variabili di separazione, per ogni categoria di ciascuna variabile di separazione verrà creato un gruppo distinto all'interno delle categorie della variabile di separazione precedente nell'elenco. Le variabili di separazione devono essere variabili categoriali discrete che suddividono i casi in un numero limitato di categorie significative. I singoli valori di ciascuna variabile di separazione vengono visualizzati, ordinati, in una colonna distinta a sinistra di tutte le colonne dati.

Report. Consente di controllare le caratteristiche generali del report, inclusi i titoli, le statistiche riassuntive globali, la visualizzazione dei valori mancanti e la numerazione delle pagine.

Visualizza casi. Consente di visualizzare i valori effettivi (o etichette dei valori) delle variabili delle colonne di dati per ciascun caso. In tal modo viene creato un listato che potrebbe risultare molto più lungo di un report riepilogativo.

Bozza. Consente di visualizzare solo la prima pagina del report. Questa opzione è utile per visualizzare in anteprima il formato del report prima di generarlo.

I dati sono già ordinati. Se il report include variabili di separazione, prima di generarlo è necessario ordinare il file dati in base ai valori delle variabili di separazione. Se il file di dati è già ordinato in base ai valori delle variabili di separazione, è possibile ridurre i tempi di elaborazione selezionando questa opzione. Questa opzione risulta particolarmente utile dopo aver visualizzato un'anteprima del report.

Esempio di output

Figura 37-1
Report combinato con elenchi dei casi e statistiche riassuntive

Reparto	Mansione per età	Mansione nella società	Valutazione	Stipendio - annuale
Moquette	27.00	3.67	2.17	$9,200
	22.00	3.92	3.08	$10,900
	23.00	3.92	3.08	$10,900
	24.00	4.00	3.25	$10,000
	30.00	4.08	3.08	$10,000
	27.00	4.33	3.17	$10,000
	33.00	2.67	2.67	$9,335
	33.00	3.75	3.25	$10,000
	44.00	4.83	4.33	$15,690
	36.00	3.83	3.25	$10,000
	35.00	3.50	3.00	$15,520
	35.00	6.00	5.33	$19,500
Media	30.75	4.04	3.31	$11,754
Apparecchi	21.00	2.67	2.67	$8,700
	26.00	2.92	2.08	$8,000
	32.00	2.92	2.92	$8,900
	33.00	3.42	2.92	$8,900
	34.00	5.08	4.50	$15,300
	24.00	3.17	3.17	$8,975
	42.00	6.50	6.50	$18,000
	30.00	2.67	2.67	$7,500
	38.00	5.00	4.42	$28,300
Media	31.11	3.81	3.54	$12,508

Per ottenere un riepilogo: Riepiloghi per righe

- Dai menu, scegliere:
  Analizza
  Report
  Report: Riepiloghi per righe...

- Selezionare una o più variabili per Colonne dati. Per ogni variabile selezionata verrà generata una colonna nel report.

- Per i report ordinati e visualizzati in base ai sottogruppi, selezionare una o più variabili per Colonne di separazione.

- Per i report con statistiche riassuntive per i sottogruppi definiti in base alle variabili di separazione, selezionare la variabile di separazione nell'elenco Colonne di

separazione e fare clic su Riepilogo nel gruppo Colonne di separazione per specificare le misure riassuntive.

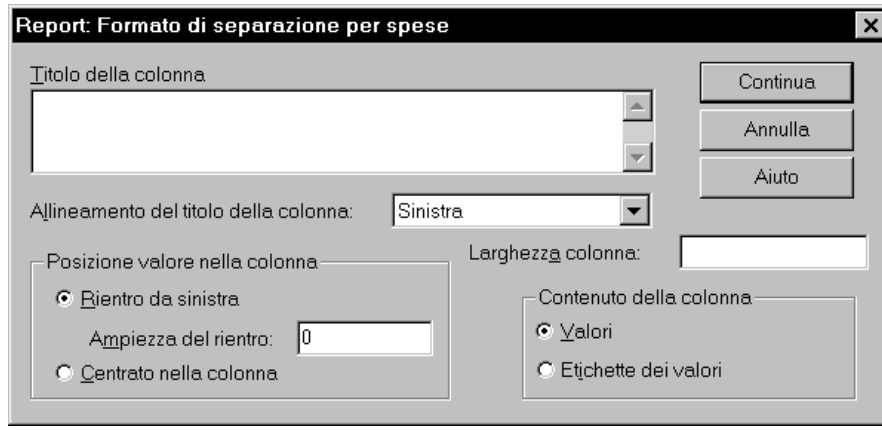
- Per i report con statistiche riassuntive globali, fare clic su Riepilogo nel gruppo Report per specificare le misure riassuntive.

Figura 37-2
Finestra di dialogo Report: Riepiloghi per righe

Formato delle colonne e di separazione del report

Nelle finestre di dialogo relative al formato è possibile impostare i titoli e la larghezza delle colonne, l'allineamento del testo e la visualizzazione dei valori e delle etichette dei valori. L'opzione Formato colonna dati consente di impostare il formato delle colonne dati nella parte destra della pagina del report. L'opzione Formato di separazione consente di impostare il formato delle colonne di separazione nella parte sinistra.

Figura 37-3
Finestra di dialogo Report: Formato colonna dati



Titolo della colonna. Consente di impostare il titolo della colonna per la variabile selezionata. I titoli lunghi vanno a capo automaticamente all'interno della colonna. Per inserire manualmente le separazioni di riga nella posizione in cui si desidera che i titoli vadano a capo, è possibile utilizzare il tasto Invio.

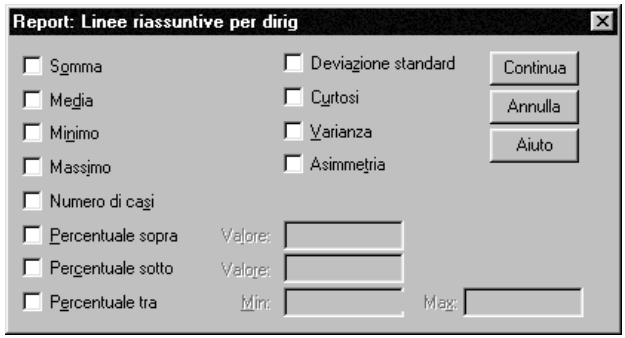
Posizione valore nella colonna. Per la variabile selezionata, consente di impostare l'allineamento dei valori o delle etichette dei valori all'interno della colonna. L'allineamento dei valori o delle etichette non modifica l'allineamento delle intestazioni di colonna. È possibile rientrare il contenuto delle colonne di un numero specifico di caratteri oppure centrarlo.

Contenuto della colonna. Per la variabile selezionata, consente di impostare la visualizzazione dei valori o delle etichette dei valori definite. I valori per i quali non è stata definita alcuna etichetta vengono sempre visualizzati. Non è disponibile per le colonne dati nei report di riepilogo per colonne.

Report: Linee riassuntive per/Linee riassuntive finali

Le due finestre di dialogo Report: Linee riassuntive consentono di impostare la visualizzazione delle statistiche riassuntive per i gruppi di interruzione e per l'intero report. Linee riassuntive consente di impostare le statistiche dei sottogruppi per ciascuna categoria definita tramite le variabili di separazione. Linee riassuntive finali consente di impostare le statistiche globali visualizzate nella parte finale del report.

Figura 37-4
Finestra di dialogo Report: Linee riassuntive

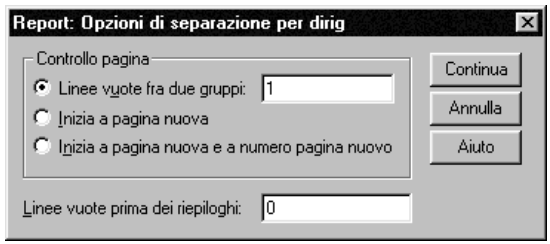


Le statistiche di riepilogo disponibili sono: somma, media, minimo, massimo, numero di casi, percentuale di casi al di sopra o al di sotto di un valore specifico, percentuale di casi entro un intervallo specifico di valori, deviazione standard, curtosi, varianza e asimmetria.

Report: Opzioni di separazione

La funzione Opzioni di separazione consente di impostare la spaziatura e l'impaginazione delle informazioni sui gruppi.

Figura 37-5
Finestra di dialogo Report: Opzioni di separazione



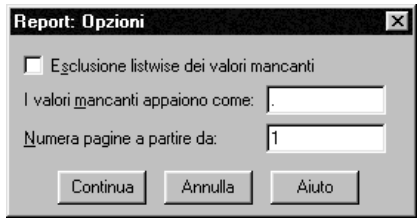
Controllo pagina. Consente di impostare la spaziatura e l'impaginazione per le categorie relative alla variabile di separazione selezionata. È possibile impostare il numero desiderato di righe vuote tra i gruppi o fare in modo che ciascun gruppo inizi in una nuova pagina.

Linee vuote prima dei riepiloghi. Consente di impostare il numero delle linee vuote tra le etichette dell'asse delle categorie o i dati e le statistiche riassuntive. Questa funzionalità risulta particolarmente utile per i report combinati che includono sia l'elenco dei singoli casi che le statistiche riassuntive per i gruppi. In questo tipo di report è possibile inserire una spaziatura tra l'elenco dei casi e le statistiche riassuntive.

Report: Opzioni

La funzione Report: Opzioni consente di impostare la modalità di elaborazione e la visualizzazione dei valori mancanti e la numerazione delle pagine del report.

Figura 37-6
Finestra di dialogo Report: Opzioni



Esclusione listwise dei valori mancanti. Consente di eliminare dal report i casi con valori mancanti per le variabili del report.

I valori mancanti appaiono come. Consente di specificare il simbolo che rappresenta i valori mancanti nel file dati. Il simbolo può essere rappresentanto da un solo carattere, ma può essere usato sia per i valori **mancanti di sistema** che **definiti dall'utente**.

Numera pagine a partire da. Consente di specificare un numero di pagina con cui contrassegnare la prima pagina del report.

Report: Layout

L'opzione Report: Layout consente di impostare la larghezza e la lunghezza di ogni pagina del report, la posizione del report nella pagina e l'inserimento di linee vuote ed etichette.

Figura 37-7
Finestra di dialogo Report: Layout

Layout di pagina. Consente di impostare i margini della pagina espressi in linee (superiori e inferiori) e caratteri (a destra e a sinistra) nonché l'allineamento del report all'interno dei margini.

Titoli e piè di pagina. Consente di impostare il numero delle linee che separano i piè di pagina e i titoli della pagina dal corpo del report.

Colonne di separazione. Consente di impostare la visualizzazione delle colonne di separazione. Se vengono specificate più variabili di separazione, queste possono trovarsi in colonne diverse oppure nella prima colonna. Se tutte le variabili di separazione vengono inserite nella prima colonna, verrà creato un report di larghezza minore.

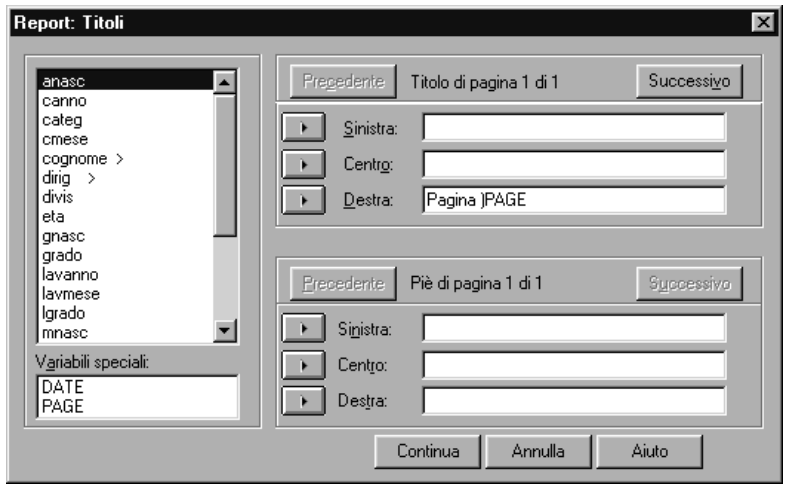
Titoli di colonna. Consente di impostare la visualizzazione dei titoli di colonna, inclusi lo spazio tra i titoli e il corpo del report, la sottolineatura del titolo e l'allineamento verticale dei titoli di colonna.

Spaziature ed etichette di separazione. Consente di posizionare le informazioni relative alle colonne dati (valori e/o statistiche riassuntive) in relazione alle etichette di separazione all'inizio di ogni gruppo. La prima riga delle informazioni sulla colonna dati può iniziare sulla stessa riga in cui si trova l'etichetta di gruppo o dopo un numero specifico di righe rispetto alla posizione dell'etichetta di gruppo. Non è disponibile per i report di riepilogo per colonne.

Report: Titoli

L'opzione Report: Titoli consente di impostare il contenuto e la posizione dei titoli e dei piè di pagina dei report. È possibile specificare fino a dieci righe per i titoli di pagina e per i piè di pagina, con componenti centrati oppure allineati a destra o a sinistra in ciascuna riga.

Figura 37-8
Finestra di dialogo Report: Titoli



Se si inseriscono variabili in titoli e in piè di pagina, l'etichetta dei valori corrente o il valore della variabile verrà visualizzato nel titolo o nella nota a piè di pagina. Nei titoli viene visualizzata l'etichetta dei valori corrispondente al valore della variabile all'inizio della pagina. Nei piè di pagina viene visualizzata l'etichetta dei valori corrispondente al valore della variabile alla fine della pagina. Se non sono presenti etichette dei valori, viene visualizzato il valore effettivo.

Variabili speciali. Le variabili speciali *DATE* e *PAGE* consentono di inserire la data corrente o il numero di pagina in una delle righe dell'intestazione o del piè di pagina del report. Se il file dati utilizzato contiene le variabili denominate *DATE* o *PAGE*, non sarà possibile utilizzare tali variabili nei titoli o nei piè di pagina.

Report: Riepiloghi per colonne

L'opzione Report: Riepiloghi per colonne consente di creare report di riepilogo in cui le statistiche riassuntive vengono visualizzate in colonne distinte.

Esempio. Una ditta proprietaria di una catena di negozi al dettaglio registra le informazioni sui dipendenti. Tra queste sono comprese informazioni sugli stipendi e le mansioni nonché sul reparto in cui lavora ogni dipendente. È quindi possibile creare un report che includa le statistiche riassuntive sugli stipendi (ad esempio media, minimo e massimo) per ogni reparto.

Colonne dati. Consente di visualizzare un elenco delle variabili da rappresentare nel report e per le quali si desidera produrre statistiche riassuntive e di impostare il formato di visualizzazione e le statistiche riassuntive visualizzate per ogni variabile.

Colonne di separazione. Consente di visualizzare l'elenco delle variabili di separazione facoltative che suddividono il report in più gruppi e di impostare i formati di visualizzazione delle colonne di separazione. Se sono presenti più variabili di separazione, per ogni categoria di ciascuna variabile di separazione verrà creato un gruppo distinto all'interno delle categorie della variabile di separazione precedente nell'elenco. Le variabili di separazione devono essere variabili categoriali discrete che suddividono i casi in un numero limitato di categorie significative.

Report. Consente di impostare le caratteristiche globali del report, inclusi i titoli, la visualizzazione dei valori mancanti e la numerazione delle pagine.

Bozza. Consente di visualizzare solo la prima pagina del report. Questa opzione è utile per visualizzare in anteprima il formato del report prima di generarlo.

I dati sono già ordinati. Se il report include variabili di separazione, prima di generarlo è necessario ordinare il file dati in base ai valori delle variabili di separazione. Se il file di dati è già ordinato in base ai valori delle variabili di separazione, è possibile ridurre i tempi di elaborazione selezionando questa opzione. Questa opzione risulta particolarmente utile dopo aver visualizzato un'anteprima del report.

Esempio di output

Figura 37-9
Report di riepilogo con le statistiche riassuntive nelle colonne

Reparto	Età annuale media	Stipendio annuale minimo	Stipendio annuale massimo	Stipendio
Moquette	30.75	$11,754	$9,200	$19,500
Apparecchi	31.11	$12,508	$7,500	$28,300
Mobilio	36.87	$13,255	$8,975	$17,050
Hardware	36.20	$17,580	$7,450	$22,500

Per ottenere un riepilogo: Riepiloghi per colonne

- Dai menu, scegliere:
  Analizza
  Report
  Report: Riepiloghi per colonne...

- Selezionare una o più variabili per Colonne dati. Per ogni variabile selezionata verrà generata una colonna nel report.

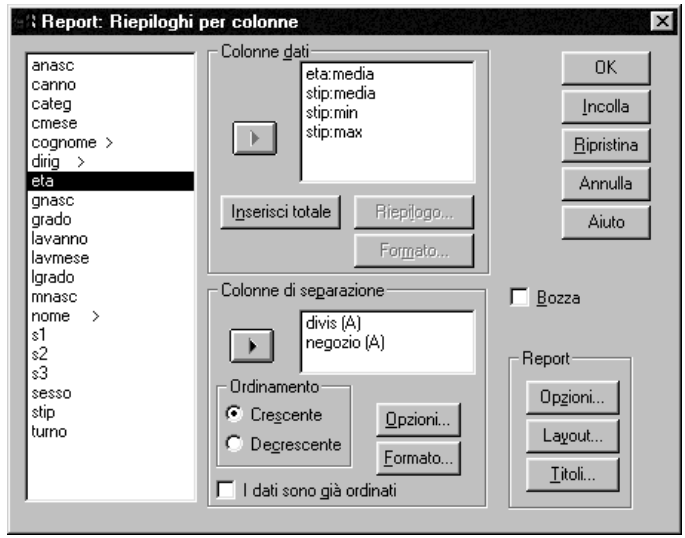
- Per modificare la misura riassuntiva relativa a una variabile, selezionare la variabile nell'elenco Colonne dati e fare clic su Riepilogo.

- Per ottenere più di una misura riassuntiva per una variabile, selezionare la variabile nell'elenco sorgente e spostarla nell'elenco Colonne dati più volte, una per ogni misura riassuntiva desiderata.

- Per visualizzare una colonna contenente la somma, la media, il rapporto o altre funzioni per le colonne esistenti, fare clic su Inserisci totale. In tal modo verrà inserita una variabile denominata *totale* nell'elenco Colonne dati.

- Per i report ordinati e visualizzati in base ai sottogruppi, selezionare una o più variabili per Colonne di separazione.

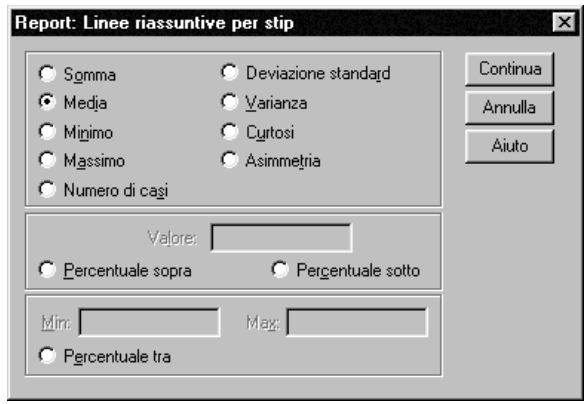
Figura 37-10
Finestra di dialogo Report: Riepiloghi per colonne



Funzione di rappresentazione delle colonne di dati

L'opzione Linee riassuntive consente di controllare le statistiche riassuntive visualizzate per la variabile della colonna dati selezionata.

Figura 37-11
Finestra di dialogo Report: Linee riassuntive



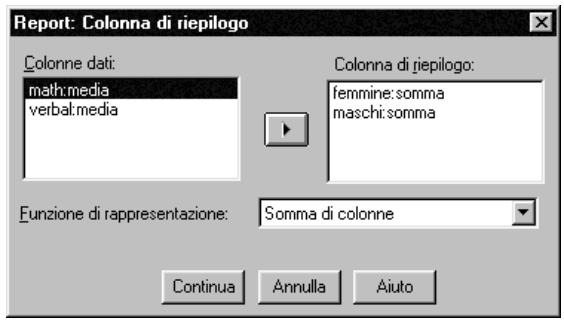
Le statistiche di riepilogo disponibili sono: somma, media, minimo, massimo, numero di casi, percentuale di casi al di sopra o al di sotto di un valore specifico, percentuale di casi entro un intervallo specifico di valori, deviazione standard, varianza, curtosi e asimmetria.

Colonna di riepilogo del totale generale

L'opzione Colonna di riepilogo consente di gestire le statistiche riassuntive generali che riassumono due o più colonne dati.

Le statistiche riassuntive generali disponibili sono somma di colonne, media di colonne, minimo, massimo, differenza tra valori in due colonne, quoziente dei valori in una colonna divisi per i valori in un'altra colonna e prodotto di valori di colonne moltiplicati insieme.

Figura 37-12
Finestra di dialogo Report: Colonna di riepilogo



Somma di colonne. La colonna *totale* rappresenta la somma delle colonne nell'elenco Colonna di riepilogo.

Media di colonne. La colonna *totale* rappresenta la media delle colonne nell'elenco Colonna di riepilogo.

Minimo di colonne. La colonna *totale* rappresenta il minimo delle colonne nell'elenco Colonna di riepilogo.

Massimo di colonne. La colonna *totale* rappresenta il massimo delle colonne nell'elenco Colonna di riepilogo.

1a colonna – 2a colonna. La colonna *totale* rappresenta la differenza delle colonne nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

1a colonna / 2a colonna. La colonna *totale* rappresenta il quoziente delle colonne nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

% 1a colonna / 2a colonna. La colonna *totale* rappresenta la percentuale della prima colonna rispetto alla seconda colonna nell'elenco Colonna di riepilogo. L'elenco Colonna di riepilogo deve contenere esattamente due colonne.

Prodotto di colonne. La colonna *totale* rappresenta il prodotto delle colonne nell'elenco Colonna di riepilogo.

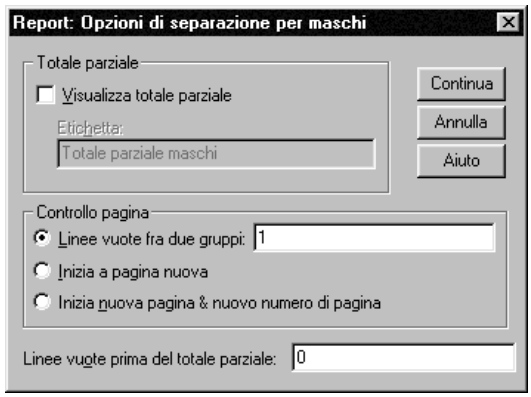
Formato delle colonne del report

Le opzioni di formattazione delle colonne dati e di separazione per i report di riepilogo per colonne sono le stesse descritte per i report di riepilogo per righe.

Report: Opzioni di separazione (Riepiloghi per colonne)

La funzione Opzioni di separazione consente di impostare la visualizzazione, la spaziatura e l'impaginazione per i gruppi.

Figura 37-13
Finestra di dialogo Report: Opzioni di separazione



Totale parziale. Consente di impostare la visualizzazione di totali parziali per i gruppi.

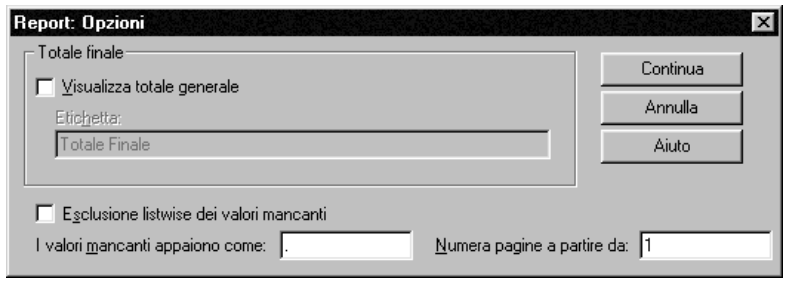
Controllo pagina. Consente di impostare la spaziatura e l'impaginazione per le categorie relative alla variabile di separazione selezionata. È possibile impostare il numero desiderato di righe vuote tra i gruppi o fare in modo che ciascun gruppo inizi in una nuova pagina.

Linee vuote prima del totale parziale. Consente di impostare il numero di righe vuote tra i dati dei gruppi e i totali parziali.

Report: Opzioni (Riepiloghi per colonne)

Le opzioni consentono di impostare la visualizzazione dei totali generali e dei valori mancanti e l'impaginazione dei report di riepilogo per colonne.

Figura 37-14
Finestra di dialogo Report: Opzioni



Totale finale. Consente di visualizzare ed etichettare un totale finale per ogni colonna, visualizzato alla fine della colonna.

Valori mancanti. È possibile escludere i valori mancanti dal report o selezionare un solo carattere per indicare i valori mancanti nel report.

Report: Layout per riepiloghi per colonne

Le opzioni di layout per i report di riepilogo per colonne sono le stesse descritte per i report di riepilogo per righe.

Funzioni aggiuntive del comando REPORT

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Visualizzare diverse funzioni di riepilogo nelle colonne di una linea riassuntiva.
- Inserire linee riassuntive nelle colonne dati per le variabili diverse dalle variabili della colonna, o per le varie combinazioni (funzioni composte) di funzioni di rappresentazione.
- Utilizzare Mediana, Moda, Frequenza e Percentuale come funzioni di rappresentazione.
- Controllare in modo più preciso il formato di visualizzazione delle statistiche riassuntive.
- Inserire linee vuote in corrispondenza di vari punti nei report.
- Inserire linee vuote dopo ogni *n* caso nei listati.

A causa della complessità della sintassi `REPORT`, può risultare utile, durante la costruzione di un nuovo report con la sintassi, approssimare il report generato dalle finestre di dialogo, copiare e incollare la sintassi corrispondente e ridefinire tale sintassi in modo da farla corrispondere al report specifico.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
38

Analisi di affidabilità

L'analisi di affidabilità consente di studiare le proprietà delle scale di misurazione e degli elementi che le compongono. La procedura Analisi di affidabilità calcola una serie di misure comunemente utilizzate in relazione all'affidabilità della scala e fornisce inoltre informazioni relative alle relazioni tra singoli elementi della scala. I coefficienti di correlazione tra classi possono essere utilizzati per calcolare le stime di affidabilità.

Esempio. Il questionario misura la soddisfazione del cliente in un modo utile? Utilizzando l'analisi di affidabilità, è possibile determinare il grado di correlazione tra gli elementi del questionario, ottenere un indice globale della ripetibilità oppure la concordanza interna della scala in modo globale. È quindi possibile identificare gli elementi del problema che devono essere esclusi dalla scala.

Statistiche. Descrittive per ogni variabile e per la scala, statistiche riassuntive degli elementi, correlazioni e covarianze tra elementi, stime di affidabilità, tabella ANOVA, coefficienti di correlazione tra classi, *T*quadratodi Hotelling e test di additività di Tukey.

Modelli. Sono disponibili i seguenti modelli di affidabilità:

- **Alfa (Cronbach).** È un modello di concordanza interna, basato sulla media di correlazione fra elementi.
- **Divisione a metà.** Questo modello divide la scala in due parti ed esamina la correlazione tra le parti.
- **Guttman.** Questo modello calcola i limiti inferiori di Guttman per una reale affidabilità.
- **Parallelo.** Questo modello presume che tutti gli elementi abbiano varianze e varianze di errore uguali tra le replicazioni.
- **Parallelo esatto.** Questo modello afferma le ipotesi del modello parallelo e assume inoltre medie uguali degli elementi.

Dati. I dati possono essere dicotomici, ordinali oppure intervalli ma devono essere codificati numericamente.

Assunzioni. Le osservazioni devono essere indipendenti e gli errori non devono essere correlati agli elementi. Ogni coppia di elementi deve avere una distribuzione normale bivariata. Le scale devono essere additive, in modo che ogni elemento sia correlato in modo lineare al punteggio totale.

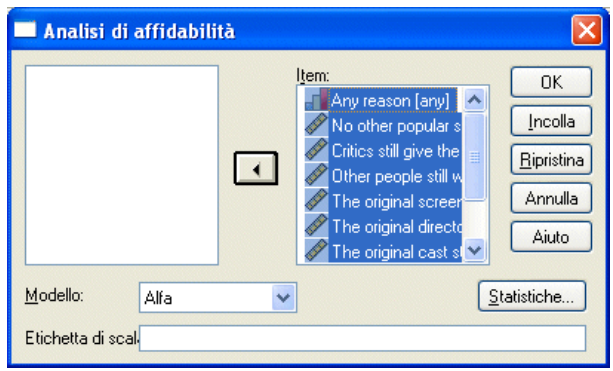
Procedure correlate. Se si desidera esplorare la dimensionalità degli elementi della scala (per verificare se è necessaria più di una costruzione per tenere conto del modello dei punteggi degli elementi), utilizzare la procedura Analisi fattoriale o Scaling multidimensionale. Per identificare i gruppi omogenei di variabili, usare la cluster gerarchica per raggruppare le variabili.

Per ottenere l'analisi di affidabilità

- Dai menu, scegliere:

Analizza
  Scala
    Analisi di affidabilità...

Figura 38-1
Finestra di dialogo Analisi di affidabilità

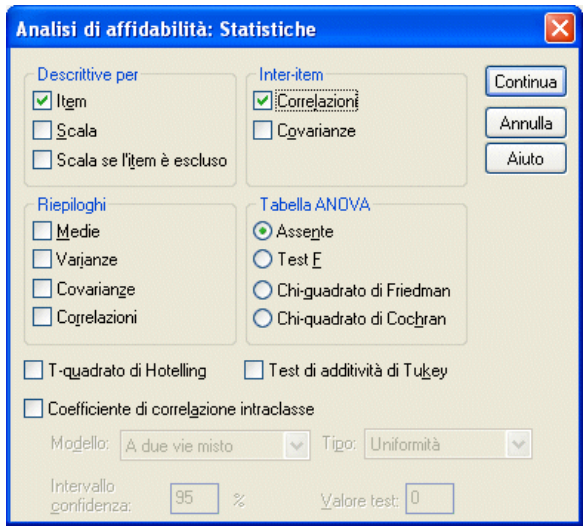


- Selezionare due o più variabili come potenziali componenti di una scala additiva.

- Scegliere un modello dall'elenco a discesa Modello.

Analisi di affidabilità: Statistiche

Figura 38-2
Finestra di dialogo Analisi di affidabilità: Statistiche



È possibile selezionare varie statistiche per la descrizione della scala e degli elementi. Le statistiche riportate per impostazione predefinita comprendono il numero di casi, il numero di elementi e le stime di affidabilità riportati di seguito:

- Modelli Alfa: coefficiente alfa. Per i dati dicotomici, è equivalente al coefficiente Kuder-Richardson 20 (KR20).
- Modelli Divisione a metà: correlazione tra stime dei parametri, affidabilità di Divisione a metà di Guttman, affidabilità di Spearman-Brown (lunghezza uguale e diversa) e coefficiente alfa per ogni metà.
- Modelli di Guttman: coefficienti di affidabilità da lambda 1 a lambda 6.
- Modelli Parallelo e Parallelo esatto: test sulla bontà dell'adattamento del modello, stime della varianza di errore, matrice di correlazione e troncamento, media comune stimata, affidabilità stimata e dati non troncati.

Descrittive per. Fornisce statistiche descrittive per le scale o per gli elementi tra i casi.

- **Elemento.** Fornisce statistiche descrittive per gli elementi tra i casi.

- **Scala.** Fornisce statistiche descrittive per le scale.
- **Scala se l'item è escluso.** Consente di visualizzare statistiche riassuntive per il confronto di ogni elemento con la scala composta dagli altri elementi. Le statistiche includono la media della scala e la varianza risultante se l'elemento venisse eliminato dalla scala, la correlazione tra l'elemento e la scala composta dagli altri elementi e l'Alfa di Cronbach risultante se l'elemento venisse eliminato dalla scala.

Riepiloghi. Fornisce statistiche descrittive della distribuzione di elementi tra tutti gli elementi nella scala.

- **Medie (Affidabilità).** Produce statistiche riassuntive basate sulle medie degli item. Vengono visualizzate la media più piccola, la più grande e la media centrale, nonché l'intervallo e la deviazione standard delle medie e il rapporto fra la media più grande e la più piccola.
- **Varianze.** Statistiche riassuntive per le varianze degli elementi. Vengono riprodotte la varianza minima, massima e media, nonché l'intervallo e il rapporto fra varianza massima e minima.
- **Covarianze.** Produce statistiche riassuntive basate sulle covarianze fra item. Vengono visualizzate la covarianza più piccola, la più grande e la covarianza media, nonché l'intervallo e la deviazione standard delle covarianze e il rapporto fra la covarianza più grande e la più piccola.
- **Correlazioni.** Produce statistiche riassuntive basate sulle correlazioni fra item. Vengono visualizzate la correlazione più piccola, la più grande e la correlazione media, nonché l'intervallo e la deviazione standard delle correlazioni e il rapporto fra la correlazione più grande e la più piccola.

Inter-item. Fornisce matrici di correlazioni o covarianze tra elementi.

Tabella ANOVA. Fornisce test di medie uguali.

- **Test F (Affidabilità).** Visualizza una tabella di analisi della varianza a misure ripetute.

- **Chi-quadrato di Friedman.** Visualizza il chi-quadrato di Friedman e il coefficiente W di Kendall. Questa opzione è appropriata per dati che rappresentano classifiche (ranghi). Il test chi-quadrato sostituisce il test F solitamente utilizzato nelle tabelle ANOVA.
- **Chi-quadrato di Cochran.** Visualizza la Q di Cochran. Questa opzione è appropriata per dati binari. La Q di Cochran sostituisce il test F solitamente utilizzato nelle tabelle ANOVA.

T quadrato di Hotelling. Crea un test multivariato dell'ipotesi nulla in base alla quale tutti gli elementi sulla scala hanno la stessa media.

Test di additività di Tukey. Crea un test dell'ipotesi in base alla quale non esiste un'interazione moltiplicativa tra gli elementi.

Coefficiente di correlazione intraclasse. Crea misurazioni della consistenza o dell'accordo dei valori all'interno dei casi.

- **Modello.** Consente di selezionare il modello per calcolare i coefficienti di correlazione intraclasse. I modelli disponibili sono A due vie misto, A due vie casuale e A una via casuale. Selezionare A due vie misto quando gli effetti relativi alle persone sono casuali e quelli relativi all'elemento sono fissi o A due vie casuale quando sia gli effetti relativi alle persone che quelli relativi all'elemento sono casuali. Selezionare A una via casuale quando gli effetti relativi alle persone sono casuali.
- **Tipo.** Consente di selezionare il tipo di indice. I tipi disponibili sono Uniformità e Concordanza assoluta.
- **Intervallo di confidenza.** Consente di specificare il livello relativo all'intervallo di confidenza. Il valore di default è il 95%.
- **Valore test.** Consente di specificare il valore ipotizzato del coefficiente relativo al test dell'ipotesi. Si tratta del valore rispetto al quale viene confrontato il valore osservato. Il valore di default è 0.

Opzioni aggiuntive del comando RELIABILITY

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Leggere ed analizzare una matrice di correlazione.
- Scrivere una matrice di correlazione per analizzarla in seguito.
- Specificare le divisioni diverse dalle metà uguali per il metodo di divisione a metà.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
39

Scaling multidimensionale

La procedura Scaling multidimensionale consente di effettuare un tentativo per trovare la struttura in un insieme di misure di distanza tra oggetti o casi. Questa operazione viene compiuta assegnando le osservazioni a posizioni specifiche in uno spazio concettuale (in genere bi o tridimensionale) in modo che le distanze tra i punti nello spazio corrispondano il più possibile alle dissimilarità specificate. In molti casi, le dimensioni di questo spazio concettuale possono essere interpretate ed utilizzate allo scopo di comprendere meglio i dati.

Se si dispone di variabili misurate oggettivamente, è possibile utilizzare lo scaling multidimensionale come una tecnica di riduzione dei dati (la procedura calcolerà le distanze dai dati multivariati per l'utente, se necessario). È inoltre possibile applicare lo scaling multidimensionale alle stime soggettive di dissimilarità tra oggetti o concetti. In aggiunta, Scaling multidimensionale può gestire i dati di dissimilarità da più origini, come nel caso di più stimatori o di rispondenti ai questionari.

Esempio. In che modo i consumatori percepiscono le similitudini tra automobili diverse? Se si dispone di dati che rilevano similarità tra diverse forme e modelli di automobili, lo scaling multidimensionale consentirà di identificare le dimensioni in grado di descrivere le percezioni dei consumatori. È possibile verificare, ad esempio, che il prezzo e le dimensioni di un veicolo definiscono uno spazio bidimensionale, secondo le spiegazioni fornite dai rispondenti.

Statistiche. Per ciascun modello: matrice dei dati, matrice di dati scalati ottimale, s-stress (di Young), stress (di Kruskal), RSQ, coordinate degli stimoli, media dello stress e RSQ per ogni stimolo (modelli RMDS). Per modelli di singole differenze (INDSCAL): pesi di soggetto e valore non ponderato per ogni soggetto. Per ogni matrice in modelli di scaling multidimensionale replicati: stress e RSQ per ogni stimolo. Grafici: coordinate degli stimoli (bi o tridimensionali), grafico a dispersione delle disparità rispetto alle distanze.

Dati. Se si dispone di dati di dissimilarità, tutte le dissimilarità dovrebbero essere quantitative e dovrebbero essere misurate in base alla stessa metrica. Se i dati sono multivariati, le variabili possono essere quantitative, binarie o dati di conteggio. Lo scaling delle variabili è una questione importante poiché le differenze possono influenzare la soluzione. Se le variabili hanno differenze significative (ad esempio, una variabile è misurata in dollari e l'altra è misurata in anni), è consigliabile standardizzarle. Questa operazione può essere eseguita automaticamente dalla procedura Scaling multidimensionale.

Assunzioni. La procedura Scaling multidimensionale è relativamente libera da ipotesi di distribuzione. Assicurarsi di selezionare il livello di misurazione appropriato (ordinale, intervallo o rapporto) nella finestra di dialogo Scaling multidimensionale per essere sicuri che i risultati vengano calcolati correttamente.

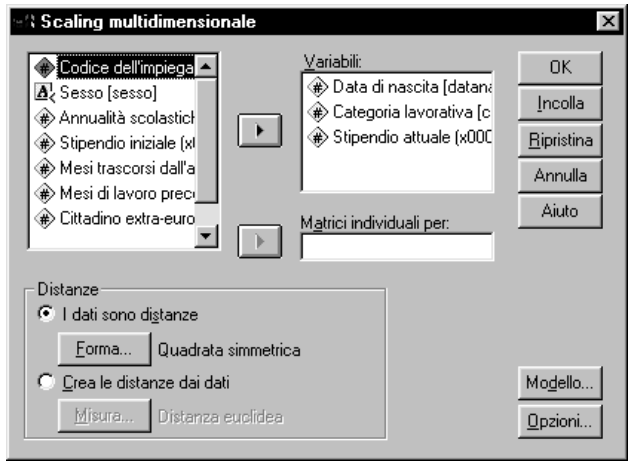
Procedure correlate. Se l'obiettivo è la riduzione dei dati, si può considerare un metodo alternativo quale l'analisi fattoriale, in particolare se le variabili sono di tipo quantitativo. Se si desidera identificare gruppi di casi simili, considerare la possibilità di integrare l'analisi di scaling multidimensionale con un'analisi gerarchica o cluster *k*-medie.

Per ottenere un'analisi Scaling multidimensionale

▶ Dai menu, scegliere:

Analizza
  Scala
    Scaling multidimensionale...

Figura 39-1
Finestra di dialogo Scaling multidimensionale



- Nel gruppo Distanze selezionare I dati sono distanze oppureCrea le distanze dai dati.
- Se i dati sono rappresentati da distanze, selezionare almeno quattro variabili numeriche per l'analisi. È possibile anche fare clic su Forma per specificare la forma della matrice delle distanze.
- Se si desidera che SPSS crei le distanze prima di analizzarle, selezionare almeno una variabile numerica. È possibile anche fare clic su Misura per specificare il tipo di misura di distanza desiderato. Ciascuna variabile di raggruppamento può essere un numero o una stringa ed è possibile creare matrici diverse per ciascuna categoria di variabili spostando la variable in questione in Matrici singole da elencare.

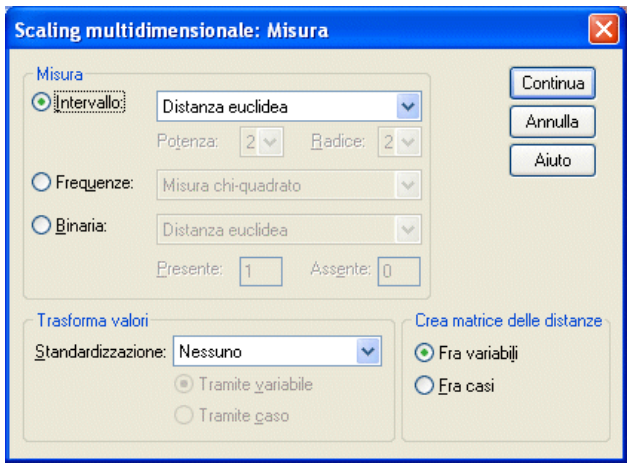
Scaling multidimensionale: Forma dei dati

Figura 39-2
Finestra di dialogo Forma scaling multidimensionale

Se il file dati attivo rappresenta le distanze tra uno o due insiemi di oggetti, è necessario specificare la forma della matrice dei dati in modo da ottenere i risultati corretti. *Nota*: Non è possibile selezionare Quadrata simmetrica se la finestra di dialogo Modello specifica la riga in modo condizionale.

Scaling multidimensionale: Crea misure dai dati

Figura 39-3
Finestra di dialogo Scaling multidimensionale: Crea misure dai dati



La procedura Scaling multidimensionale utilizza dati di dissimilarità per creare una soluzione di scaling. Se i dati disponibili sono dati multivariati (valori di variabili misurate), è necessario creare dati di dissimilarità in modo da calcolare una soluzione di scaling multidimensionale. È possibile specificare i dettagli della creazione delle misure di dissimilarità a partire dai dati disponibili.

Misura. Consente di specificare la misura di dissimilarità per l'analisi. Selezionare un'alternativa dal gruppo Misura corrispondente al tipo di dati desiderato e quindi selezionare una delle misure dall'elenco a discesa corrispondente a tale tipo di misura. Le alternative disponibili sono:

- **Intervallo**. Distanza euclidea, Distanza euclidea quadratica, Chebychev, City-Block, Minkowski o Personalizzato.

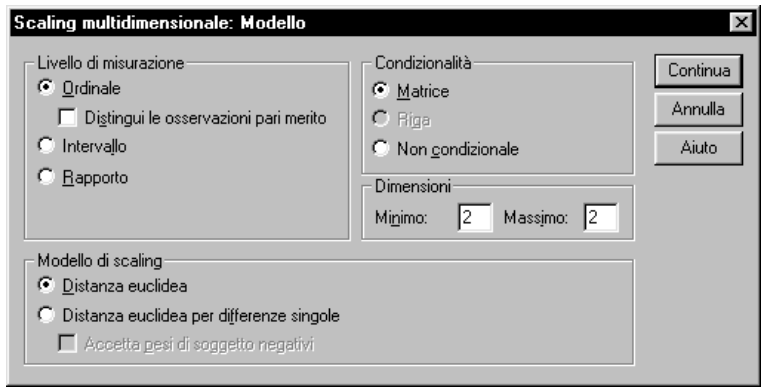
- **Conteggio**. Misura chi-quadrato e Misura phi-quadrato.
- **Binaria**. Distanza euclidea, Distanza euclidea quadratica, Differenza di dimensione, Differenza di modello, Varianza o Lance e Williams.

Crea matrice delle distanze. Consente di scegliere l'unità di analisi. Le alternative sono Fra variabili o Fra casi.

Trasforma valori. In alcuni casi, ad esempio quando le variabili sono misurate su scale molto diverse, è possibile standardizzarne i valori prima di calcolare le dissimilarità (non applicabile ai dati binari). Selezionare un metodo di standardizzazione dall'elenco a discesa Standardizza. Se non è richiesta alcuna standardizzazione, selezionare Nessuna).

Scaling multidimensionale: Modello

Figura 39-4
Finestra di dialogo Scaling multidimensionale: Modello



La stima corretta di un modello di scaling multidimensionale dipende dagli aspetti dei dati e del modello stesso.

Livello di misurazione. Consente di specificare il livello dei dati. Le alternative sono Ordinale, Intervallo o Rapporto. Se le variabili sono ordinali, la selezione dell'opzione Distingui le osservazioni pari merito richiede che vengano trattate come variabili continue, in modo che i pari merito (valori uguali per casi differenti) siano risolti in modo ottimale.

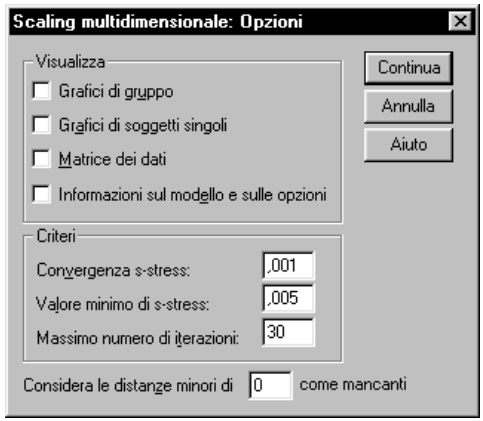
Condizionalità. Consente di specificare quali confronti sono significativi. Le alternative sono Matrice, Riga o Non condizionale.

Dimensioni. Consente di specificare la dimensione della soluzione di scaling. Per ciascun numero nell'intervallo viene calcolata una soluzione. Specificare interi da 1 a 6. È consentito un minimo di 1 solo se si è selezionato Distanza euclidea come modello di scaling. Per una soluzione singola, specificare lo stesso valore minimo e massimo.

Modello di scaling. Consente di specificare le ipotesi in base alle quali viene eseguito lo scaling. Le alternative disponibili sono Distanza euclidea o Distanza euclidea per differenze singole (nota anche come INDSCAL). Nel modello Distanza euclidea per differenze singole, è possibile selezionare l'opzione Accetta pesi di soggetto negativi, se appropriata per i dati disponibili.

Scaling multidimensionale: Opzioni

Figura 39-5
Finestra di dialogo Scaling multidimensionale: Opzioni



È possibile impostare le opzioni per l'analisi scaling multidimensionale:

Visualizzazione. Consente di selezionare vari tipi di output. Le opzioni disponibili sono Grafici di gruppo, Grafici di soggetti singoli, Matrice dei dati e Informazioni sul modello e sulle opzioni.

Criteri. Consente di determinare il momento in cui interrompere l'iterazione. Per modificare i valori di default, inserire i valori per Convergenza s-stress, Valore minimo di s-stress e Massimo numero di iterazioni.

Considera le distanze minori di N come mancanti. Le distanze minori di tale valore sono escluse dall'analisi.

Opzioni aggiuntive del comando ALSCAL

Utilizzando il linguaggio a comandi SPSS è inoltre possibile:

- Utilizzare tre tipi di modelli aggiuntivi, noti come ASCAL, AINDS e GEMSCAL, per lo scaling multidimensionale.
- Eseguire trasformazioni polinomiali su dati di intervallo e di rapporto.
- Analizzare le similarità (piuttosto che le distanze) con dati ordinali.
- Analizzare i dati nominali.
- Salvare varie matrici di coordinate e pesi nei file e rileggerli per l'analisi.
- Vincolare l'unfolding multidimensionale.

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
40

Statistiche di rapporto

La procedura Statistiche di rapporto offre un elenco completo di statistiche riassuntive per la descrizione del rapporto tra due variabili di scala.

È possibile ordinare l'output in base ai valori di una variabile di raggruppamento in ordine ascendente o discendente. È possibile eliminare il report sulle statistiche di rapporto dall'output e salvare i risultati in un file esterno.

Esempio. Esiste un buon grado di uniformità nel rapporto tra prezzo di stima e prezzo di vendita delle case in ognuno dei cinque paesi? Dall'output, si apprende che la distribuzione dei rapporti varia notevolmente da paese a paese.

Statistiche. Mediana, media, media pesata, intervalli di confidenza, coefficiente di dispersione, coefficiente di variazione centrato sulla mediana, coefficiente di variazione centrato sulla media, differenziale di prezzo, deviazione standard, deviazione assoluta media, intervallo, valori minimo e massimo e indice di concentrazione calcolato per una percentuale o un intervallo specifico per l'utente compresi nel rapporto della mediana.

Dati. Utilizzare codici numerici o stringhe corte per codificare le variabili di raggruppamento (misure di livello nominale o ordinale).

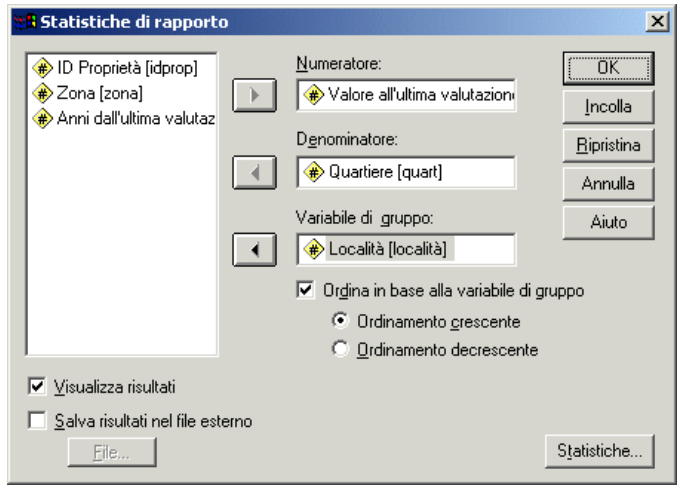
Assunzioni. Le variabili che definiscono il numeratore e il denominatore del rapporto dovrebbero essere variabili di scala con valori positivi.

Per ottenere statistiche di rapporto

▶ Dai menu, scegliere:

Analizza
  Statistiche descrittive
    Rapporto...

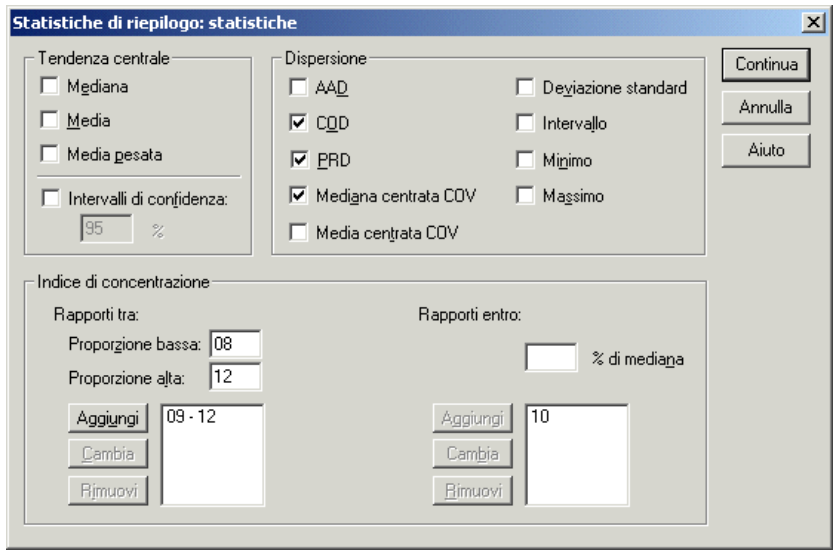
Figura 40-1
Finestra di dialogo Statistiche di rapporto



- Selezionare una variabile numeratore.
- Selezionare una variabile denominatore.
- Oppure:
  - Selezionare una variabile di raggruppamento e specificare l'ordinamento dei gruppi nei risultati.
  - Scegliere se visualizzare o meno i risultati nel Viewer.
  - Decidere se salvare o meno i risultati in un file esterno per futuri utilizzi. In caso affermativo, specificare il nome del file in cui verranno salvati.

Statistiche di rapporto

Figura 40-2
Finestra di dialogo Statistiche di rapporto



Tendenza centrale. Le misure di tendenza centrale sono statistiche che descrivono la distribuzione dei rapporti.

- **Mediana.** Il valore tale che il numero di rapporti inferiore a questo valore e il numero di rapporti maggiore di questo valore siano uguali.
- **Media.** Il risultato della somma dei rapporti e la divisione del risultato per il numero totale di rapporti.
- **Media pesata.** Il risultato della divisione della media del numeratore per la media del denominatore. La media pesata è anche la media dei rapporti ponderati dal denominatore.
- **Intervalli di confidenza.** Vengono visualizzati gli intervalli di confidenza per la media, la mediana e la media ponderata (se necessario). Specificare un valore maggiore o uguale a 0 e minore di 100 come intervallo di confidenza.

Dispersione. Queste statistiche misurano il grado di variazione o variabilità nei valori osservati.

- **AAD.** (Deviazione assoluta media) È ottenuta sommando le deviazioni assolute dei rapporti dalla mediana e dividendo il risultato per il numero totale di rapporti.
- **COD.** (Coefficiente di dispersione) È il risultato dell'espressione della deviazione assoluta media come una percentuale della mediana.
- **PRD.** (Differenziale di prezzo) Anche noto come indice di regressività, è il risultato della divisione della media per la media pesata.
- **Mediana centrata COV.** (Coefficiente di variazione) È il risultato dell'espressione delle radici quadrate medie della deviazione dalla mediana come una percentuale della mediana.
- **Media centrata COV.** (Coefficiente di variazione) È il risultato dell'espressione della deviazione standard come una percentuale della media.
- **Deviazione standard.** La deviazione standard viene ottenuta sommando le deviazioni quadrate dei rapporti dalla media, dividendo il risultato per il numero totale di rapporti meno uno ed estraendo la radice quadrata positiva.
- **Intervallo.** L'intervallo è il risultato della differenza tra rapporto massimo e rapporto minimo.
- **Minimo.** Il minimo è il rapporto più piccolo.
- **Massimo.** Il massimo è il rapporto più grande.

Indice di concentrazione. Consente di misurare la percentuale di rapporti che cade in un intervallo. È possibile calcolarlo in due modi diversi:

- **Rapporti tra.** In questo caso l'intervallo viene definito in modo esplicito specificando i valori alti e bassi dell'intervallo. Immettere i valori per le proporzioni alte e basse, quindi fare clic su Aggiungi per ottenere un intervallo.
- **Rapporti entro.** In questo caso l'intervallo viene definito in modo implicito specificando la percentuale della mediana. Digitare un valore compreso tra 0 e 100 e fare clic su Aggiungi. Il valore minimo dell'intervallo è uguale a (1 – 0,01 × valore) × mediana, e il massimo è uguale a (1 + 0,01 × valore) × mediana.

Capitolo
41

Panoramica dello strumento Grafico

Per creare grafici ad alta risoluzione è possibile utilizzare le procedure disponibili nel menu Grafici e numerose procedure del menu Analizza. In questo capitolo viene fornita una panoramica dello strumento Grafico.

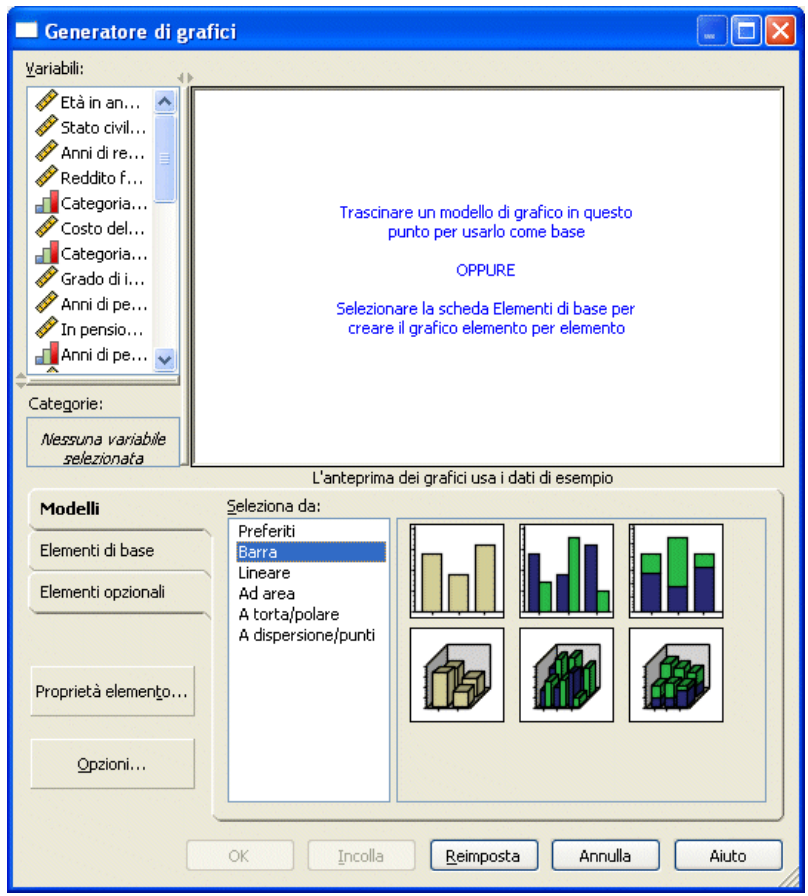
Creazione e modifica di un grafico

Prima di poter creare un grafico è necessario che i dati siano presenti nell'Editor dei dati. È possibile immettere i dati direttamente nell'Editor dei dati, aprire un file di dati salvato in precedenza, oppure leggere un foglio elettronico, un file di dati delimitato da tabulazioni o un file di database. Tramite il comando Esercitazione del menu ? è possibile accedere a esempi in linea relativi alla creazione e alla modifica di un grafico e nell'Aiuto sono disponibili informazioni sulla creazione e la modifica di tutti i tipi di grafico.

Per creare il grafico

- Dopo aver inserito i dati nell'Editor dei dati, selezionare Generatore di grafici nel menu Grafici. Verrà visualizzata la finestra di dialogo corrispondente.

Figura 41-1
Finestra di dialogo Generatore di grafici



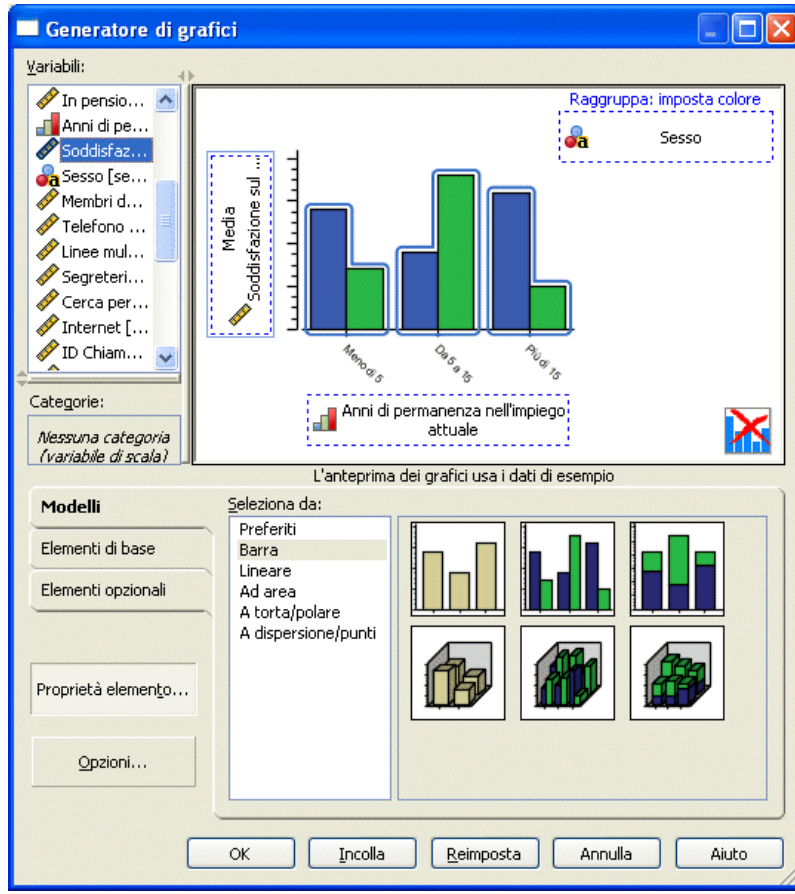
La finestra Generatore di grafici viene usata per i grafici generici visualizzati nella scheda Modelli. Per creare un grafico non disponibile nel Generatore di grafici, è possibile anche selezionare un tipo di grafico specifico nel menu Grafici.

- Nel Generatore di grafici, trascinare l'icona che rappresenta il grafico nell'area di disegno, ovvero nell'area sopra ai modelli.

- Trascinare le variabili dall'elenco Variabili all'area di inserimento dell'asse. Per visualizzare ulteriori informazioni sul Generatore di grafici, fare clic su ?.

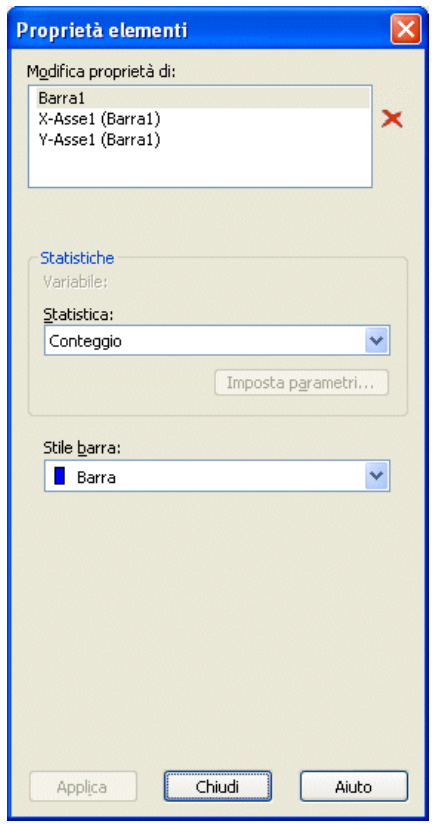
Al termine della procedura, verrà visualizzato un grafico simile al seguente.

Figura 41-2
Finestra di dialogo Generatore di grafici con aree di inserimento compilate



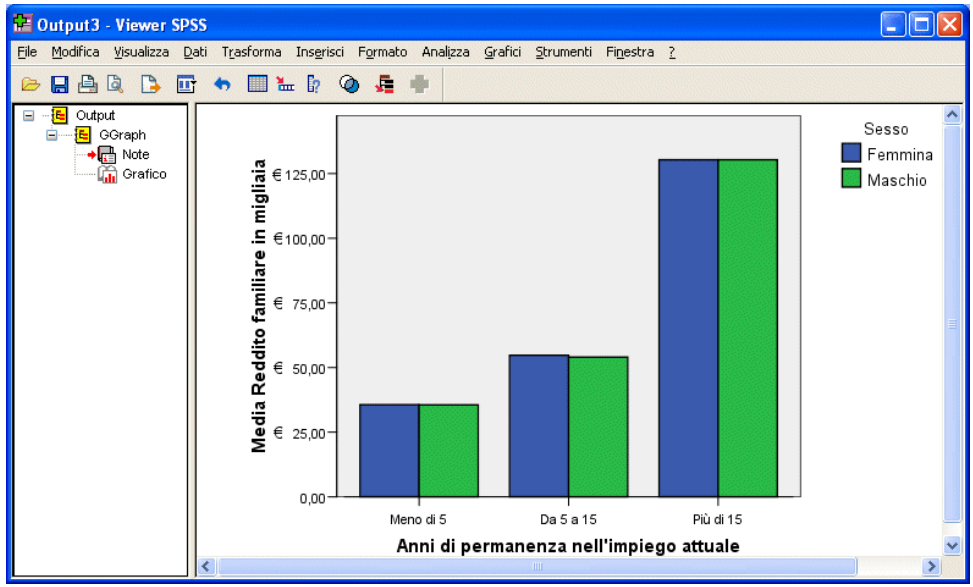
- Per modificare le statistiche o gli attributi degli assi o delle didascalie, fare clic su Proprietà Elementi.

Figura 41-3
Finestra Proprietà Elementi



- In Modifica proprietà dell'elenco, selezionare l'elemento da modificare. Per informazioni sulle proprietà specifiche, fare clic su ?.
- Fare clic su OK nella finestra di dialogo Generatore di grafici per creare il grafico. Il grafico viene visualizzato nel Viewer.

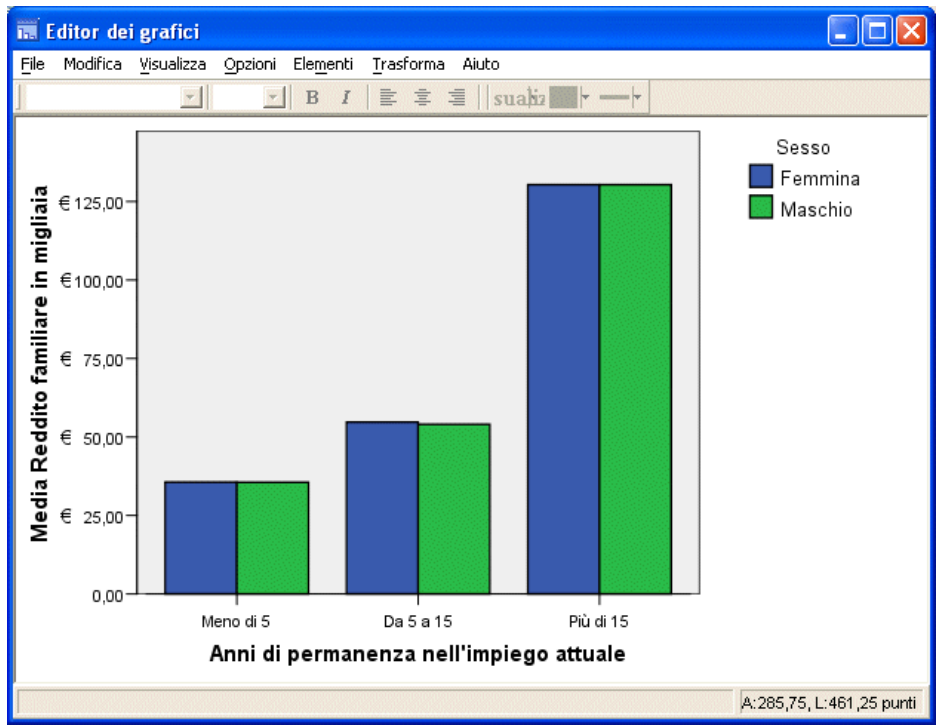
Figura 41-4
Grafico a barre visualizzato nella finestra del Viewer



Per modificare il grafico

Per modificare un grafico, fare doppio clic in un punto qualsiasi del grafico visualizzato nel Viewer. Il grafico verrà visualizzato nell'Editor dei grafici.

Figura 41-5
Grafico originale nell'Editor dei grafici



È possibile modificare qualsiasi parte del grafico o convertirlo in un tipo di grafico diverso che rappresenta gli stessi dati. È possibile aggiungere oppure visualizzare o nascondere elementi mediante i menu dell'Editor dei grafici.

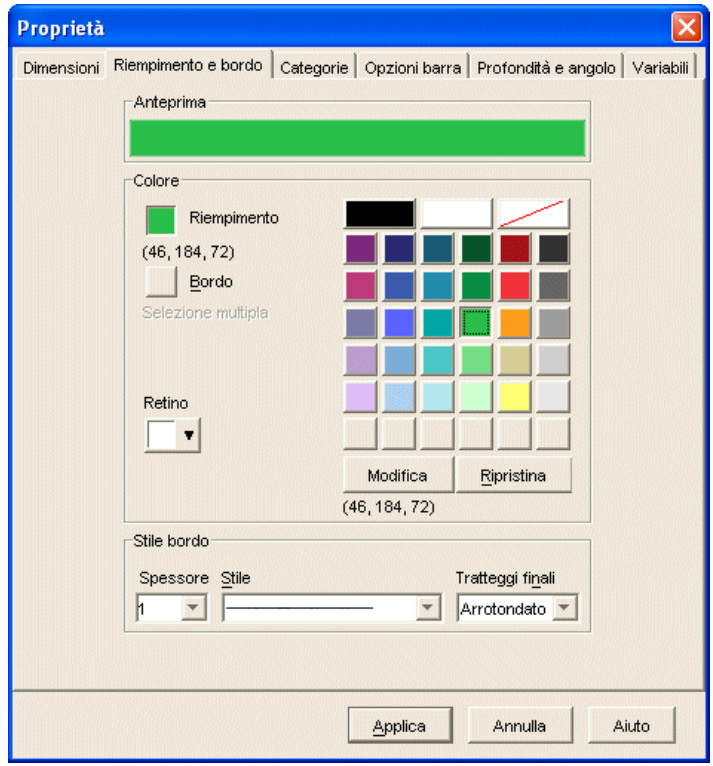
Per modificare un elemento del grafico:

- Selezionare l'elemento che si desidera modificare.
- Dai menu, scegliere:

  Modifica
  Proprietà...

  Verrà aperta la finestra Proprietà. Le schede visualizzate nella finestra Proprietà variano in base alla selezione eseguita. Per informazioni su come visualizzare le schede desiderate, vedere l'Aiuto in linea.

Figura 41-6
Finestra Proprietà

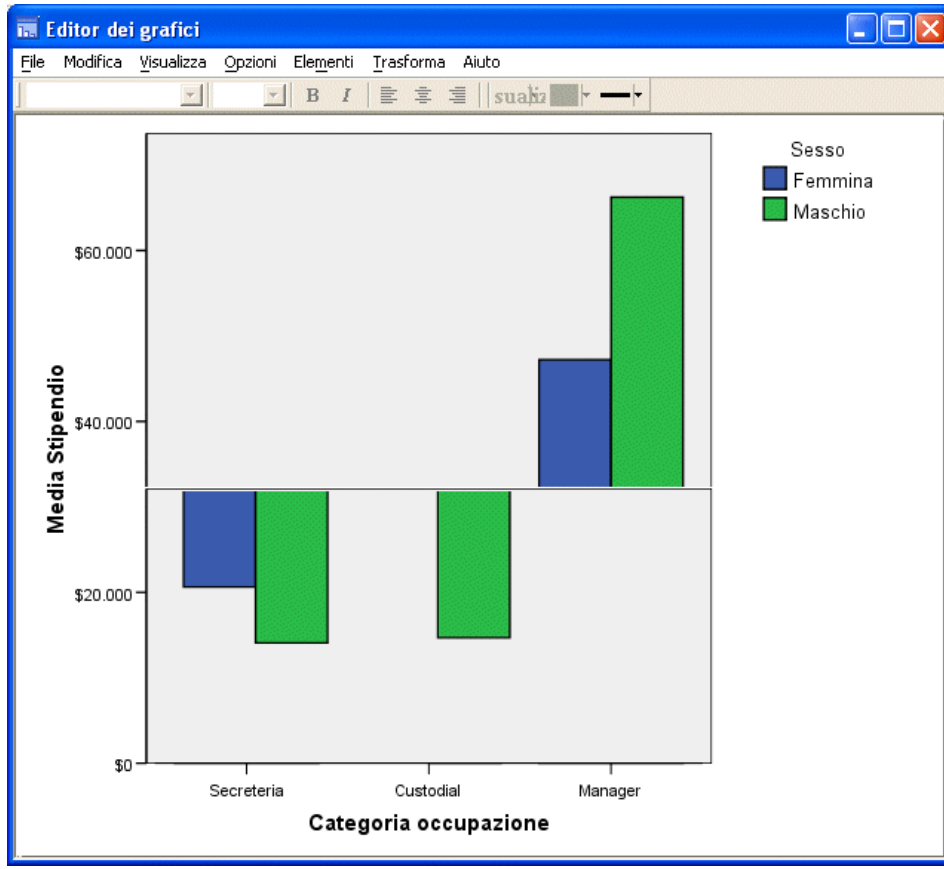


Alcune modifiche tipiche comprendono:

- Modifica del testo del grafico.
- Modifica del colore e dei retini delle barre.
- Aggiunta di testo al grafico, ad esempio aggiunta di un titolo o di un'annotazione.
- Modifica della posizione della linea di origine delle barre.

Di seguito è illustrato un grafico modificato.

Figura 41-7
Grafico modificato



Le modifiche al grafico vengono salvate quando si chiude la finestra e il grafico modificato viene visualizzato nel Viewer.

Opzioni definizione grafico

Quando si definisce un grafico, è possibile aggiungere titoli e modificare le opzioni per la creazione dei grafici. Nel Generatore di grafici, fare clic sulla scheda Elementi opzionali per specificare titoli, sottotitoli e note a piè di pagina. È possibile fare clic su Opzioni per definire le varie opzioni del grafico; ad esempio come gestire i valori

mancanti o visualizzare i modelli Nelle sezioni seguenti viene illustrato come definire queste caratteristiche durante la definizione del grafico.

Titoli, sottotitoli e note a piè di pagina

Nella definizione originale di qualsiasi grafico è possibile impostare due righe per il titolo, una riga per il sottotitolo e due righe per la nota a piè di pagina. Per specificare titoli o note a piè di pagina durante la definizione di un grafico, fare clic sulla scheda Elementi opzionali nella finestra di dialogo Generatore di grafici. Quindi, trascinare uno degli elementi nell'area di disegno. Per modificare il testo associato all'elemento, fare clic su Proprietà elemento e selezionare l'elemento dalla scheda Modifica proprietà dell'elenco.

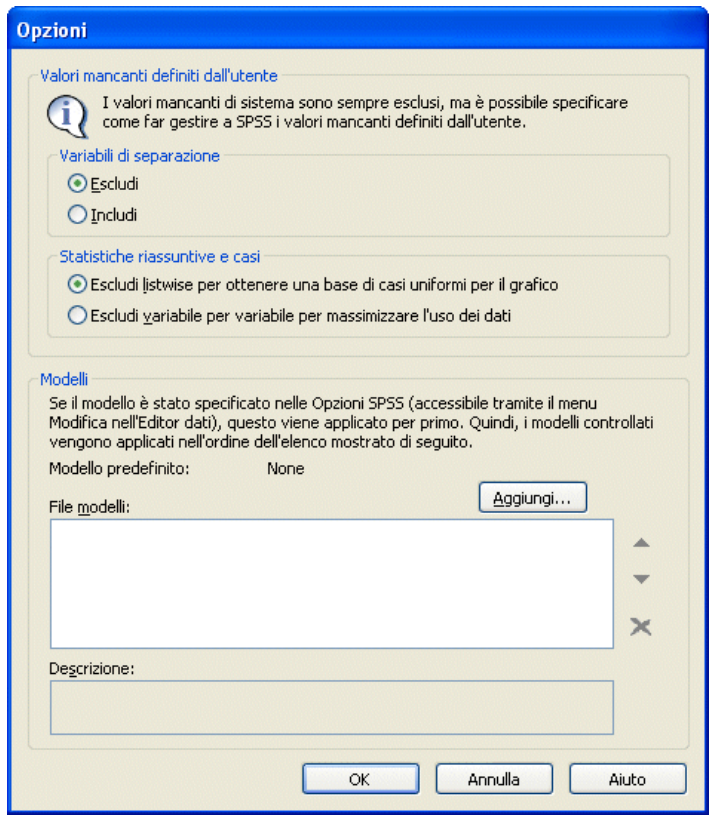
Ogni riga può contenere fino a 72 caratteri. Il numero di caratteri effetivamente inserito nel grafico varia a seconda del tipo e delle dimensioni dei caratteri. La maggior parte dei titoli è giustificata a sinistra di default e, se il titolo è troppo lungo, viene ritagliato a destra. I titoli dei grafici a torta sono allineati al centro di default e, se sono troppo lunghi, vengono ritagliati su entrambi i lati.

I titoli, i sottotitoli e le note a piè di pagina vengono resi come caselle di testo nell'Editor dei grafici. Nell'Editor dei grafici è possibile aggiungere, eliminare o aggiornare le caselle di testo, nonché modificare il tipo e la dimensione del carattere e la giustificazione.

Opzioni

La finestra di dialogo Opzioni contiene le opzioni per il grafico che si sta creando. Per visualizzare questa finestra di dialogo, fare clic su Opzioni.

Figura 41-8
Finestra di dialogo Opzioni


Opzioni
Valori mancanti definiti dall'utente
I valori mancanti di sistema sono sempre esclusi, ma è possibile specificare come far gestire a SPSS i valori mancanti definiti dall'utente.
Variabili di separazione
Escludi
Includi
Statistiche riassuntive e casi
Escludi listwise per ottenere una base di casi uniformi per il grafico
Escludi variabile per variabile per massimizzare l'uso dei dati
Modelli
Se il modello è stato specificato nelle Opzioni SPSS (accessibile tramite il menu Modifica nell'Editor dati), questo viene applicato per primo. Quindi, i modelli controllati vengono applicati nell'ordine dell'elenco mostrato di seguito.
Modello predefinito:
None
File modelli:
Aggiungi...
Descrizione:
OK
Annulla
Aiuto

Valori mancanti

Variabili di separazione

Se ci sono valori mancanti nei dati per le variabili usate per definire le categorie o i sottogruppi, selezionare Includi in modo che la categoria o le categorie dei valori mancanti definiti dall'utente (valori che sono stati identificati come mancanti dall'utente) vengano inclusi nel grafico. Queste categorie possono essere usate anche come variabili di separazione durante il calcolo di statistiche. Le categorie "mancanti" possono essere visualizzate sull'asse delel categorie o nelle didascalie, ad esempio

aggiungendo una barra aggiunta o una fetta a un grafico a torta. Se non ci sono valori mancanti, le categorie "mancanti" non vengono visualizzate.

Se si seleziona questa opzione e si desidera disattivare la visualizzazione dopo che il grafico è stato creato, selezionare il grafico e quindi scegliere Proprietà dal menu Modifica. Fare clic sulla scheda Categorie e spostare le categorie da eliminare nell'elenco Escluse.

Statistiche riepilogative e valori dei casi

È possibile selezionare una delle seguenti opzioni per escludere i casi che contengono valori mancanti:

- **Esclusione listwise per ottenere una base uniforme per il grafico.** Se qualsiasi variabile del grafico contiene un caso con un valore mancante, l'intero caso verrà escluso dal grafico.
- **Esclusione di singole variabili per ottimizzare l'uso dei dati.** Se una variabile specifica contiene casi con valori mancanti, tali casi vengono esclusi durante l'analisi della variabile.

Nelle figure seguenti vengono illustrate le differenze tra l'esclusione listwise dei casi mancanti e l'esclusione dei casi mancanti variabile per variabile. Nelle figure è riportato un grafico a barre per ciascuna delle due opzioni.

Figura 41-9
Esclusione listwise dei valori mancanti

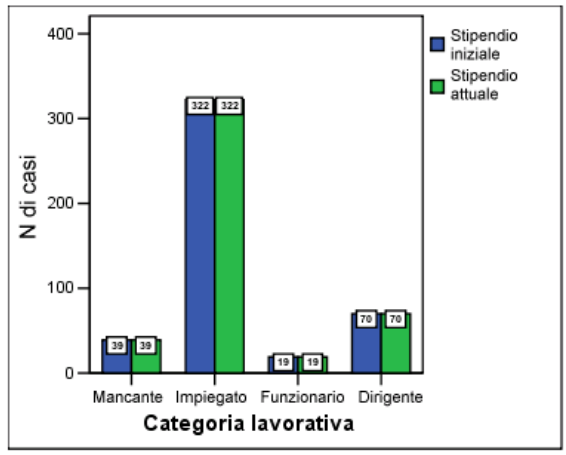
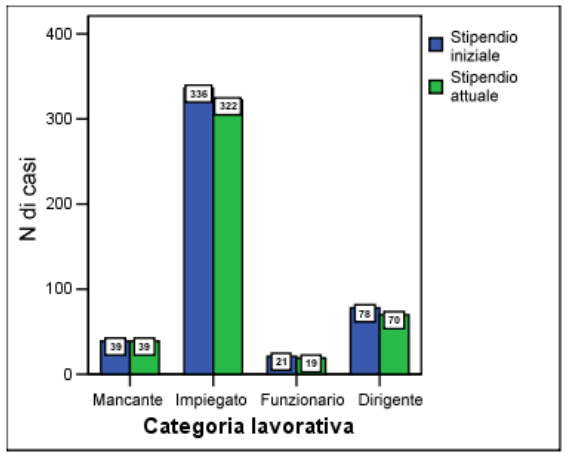


Figura 41-10
Esclusione dei valori mancanti variabile per variabile



I grafici sono stati creati utilizzando una versione del file *Employee data.sav* modificata in modo che nelle variabili relative allo stipendio attuale e alla categoria lavorativa siano presenti alcuni valori mancanti di sistema. In alcuni altri casi, è stato inserito il valore 0 che è stato definito come mancante. In entrambi i grafici è selezionata l'opzione Visualizza i gruppi definiti da valori mancanti, che aggiunge la categoria *Mancanti* alle altre categorie lavorative visualizzate. Nell etichette delle barre di ognuno dei grafici sono visualizzati i valori della funzione di rappresentazione *Numero di casi*.

In entrambi i grafici, 26 casi hanno un valore mancante di sistema per la categoria lavorativa e 13 casi hanno il valore mancante definito dall'utente (0). Nel grafico relativo all'esclusione listwise, il numero dei casi è lo stesso per entrambe le variabili in ogni raggruppamento di barre, perché se esiste un valore mancante il caso viene escluso per tutte le variabili. Nel grafico relativo all'esclusione variabile per variabile, il numero di casi non mancanti per ogni variabile viene tracciato indipendentemente dai valori mancanti delle altre variabili.

Modelli di grafico

È possibile applicare molti attributi ed elementi di testo di un grafico a un altro grafico. Questa funzione permette di modificare un grafico, salvarlo come un modello e quindi utilizzare tale modello per creare molti altri grafici simili.

Per usare un modello durante la creazione di un grafico, fare clic su Aggiungi nel gruppo Modelli. Verrà aperta un finestra di selezione dei file standard. Se si aggiungono più modelli, questi vengono applicati nell'ordine in cui appaiono nell'elenco File di modello. Tuttavia, è possibile modificare l'ordine.

Per applicare un modello a un grafico già presente nell'Editor dei grafici, dai menu scegliere:

File
  Applica modello grafico...

Verrà aperta un finestra di selezione dei file standard. Selezionare il file che si desidera utilizzare come modello. Quando si crea un nuovo grafico, il nome di file selezionato viene visualizzato nel gruppo Modello quando si torna alla finestra di dialogo per la definizione del grafico.

L'utilizzo di un modello consente di acquisire il formato di un grafico e di applicarlo al nuovo grafico che si desidera creare. Tutte le informazioni di formattazione del vecchio grafico applicabili al nuovo grafico vengono in genere applicate automaticamente. Ad esempio, se il vecchio grafico è un grafico a barre raggruppato nel quale i colori delle barre sono stati modificati in verde e giallo e il nuovo grafico è un grafico lineare multiplo, le linee di tale grafico saranno di colore giallo e verde. Se il vecchio grafico è un grafico a barre semplice con ombreggiature e il nuovo grafico è un grafico lineare semplice, le linee di tale grafico non saranno ombreggiate perché non è possibile applicare le ombreggiature ai grafici lineari. Se il modello di grafico contiene titoli che non sono presenti nel nuovo grafico, sarà possibile ottenerli dal modello. Gli eventuali titoli definiti nel nuovo grafico hanno la priorità sui titoli del modello di grafico.

Per creare un modello di grafico

- Creare un grafico.
- Inserire nel grafico gli attributi che si desidera siano presenti nel modello.
- Dai menu dell'Editor dei grafici, scegliere:

  File
    Salva modello...

- Nella finestra di dialogo Salva modello specificare le caratteristiche del grafico che si desidera salvare nel modello. Per informazioni dettagliate sulle impostazioni disponibili, vedere l'Aiuto in linea.

- Fare clic su Continua.
- Specificare un nome e un percorso per il file del nuovo modello. L'estensione del modello sarà *sgt*.

Capitolo
42

Curva ROC

Questa procedura è utile per valutare le prestazioni degli schemi di classificazione in cui i soggetti sono classificati in base a una variabile con due categorie.

Esempio. Poiché è interesse della banca classificare in modo corretto i clienti adempienti o meno, è necessario elaborare metodi specifici per decidere a chi concedere i prestiti. Le curve ROC possono essere utilizzate per valutare il livello di attendibilità di questi metodi.

Statistiche. Area sotto alla curva ROC con intervallo di confidenza e coordinate della curva ROC. Grafici: Curva ROC.

Metodi. È possibile calcolare una stima dell'area sotto alla curva ROC in modo non parametrico o parametrico mediante un modello esponenziale binegativo.

Dati. Le variabili del test sono quantitative. Le variabili del test sono spesso composte dalle probabilità dell'analisi discriminante o della regressione logistica oppure da punteggi su scala arbitraria, i quali rappresentano la "forza della convinzione" di chi classifica i soggetti in una categoria piuttosto che in un'altra. La variabile di stato può essere di qualsiasi tipo e indica la vera categoria a cui appartiene un soggetto. Il valore della variabile di stato indica quale categoria debba essere considerata *positiva*.

Assunzioni. Si suppone che i numeri crescenti sulla scala della classifica rappresentino la convinzione crescente che il soggetto appartenga a una categoria, mentre i numeri decrescenti rappresentino la convinzione crescente che il soggetto appartenga all'altra categoria. L'utente deve scegliere qual è la direzione *positiva*. Si suppone inoltre che la categoria *vero* alla quale appartiene ciascun soggetto sia conosciuta.

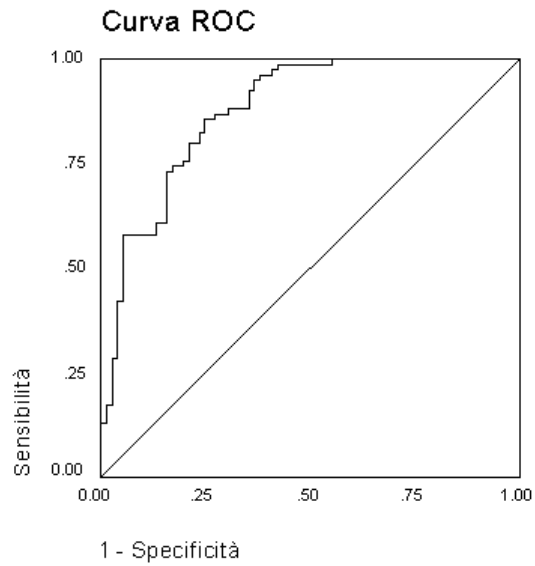
Figura 42-1
Output Curva ROC

Riepilogo dei casi

REALI	Validi (listwise)
Positivi[1]	74
Negativi	76

Più alti sono i valori delle variabili risultato del test e più evidente è lo stato reale positivo.

1. Lo stato reale positivo è 1.00



Area sottostante alla curva

Variabili risultato del test: PROBS

Area	Errore std[1]	Sig. asintotica[2]	Intervallo di confidenza asintotico al 95%	
			Limite inferiore	Limite superiore
.877	.028	.000	.823	.931

1. In base all'ipotesi non parametrica

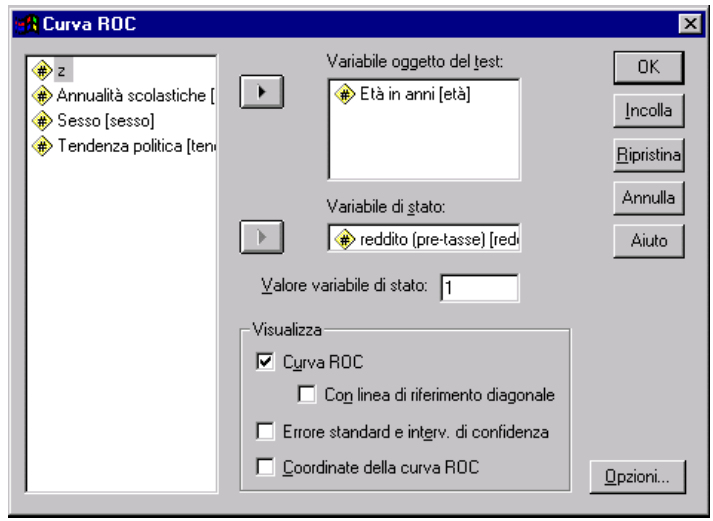
2. Ipotesi nulla: area reale = 0.5

Per creare una Curva ROC

▶ Dai menu, scegliere:

Grafici
Curva ROC...

Figura 42-2
Finestra di dialogo Curva ROC



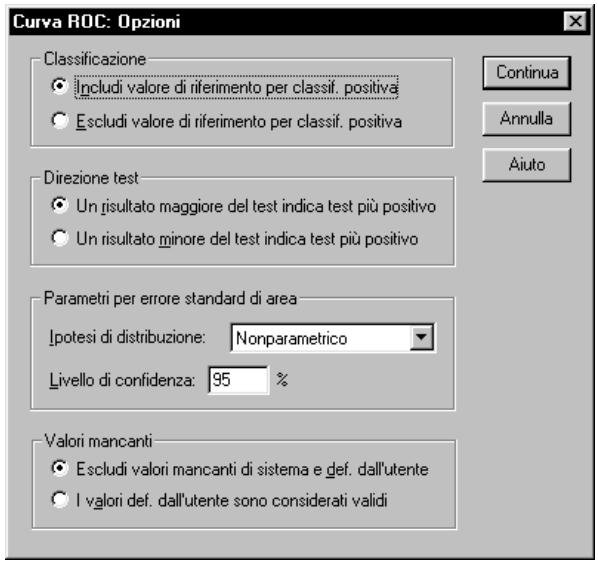
▶ Selezionare una o più variabili di probabilità per il test.

▶ Selezionare una variabile di stato.

- Individuare il valore *positivo* per la variabile di stato.

Curva ROC: Opzioni

Figura 42-3
Finestra di dialogo Curva ROC: Opzioni



È possibile selezionare le seguenti opzioni per l'analisi ROC:

Classificazione. Consente di specificare se il valore di riferimento debba essere incluso o escluso dalla classificazione *positiva*. Questa impostazione non ha attualmente alcun effetto sull'output.

Direzione test. Consente di specificare la direzione della scala in rapporto alla categoria *positiva*.

Parametri per errore standard di area. Consente di specificare il metodo di valutazione dell'errore standard dell'area sotto alla curva. I metodi disponibili sono non parametrico ed esponenziale binegativo. Consente inoltre di impostare il livello dell'intervallo di confidenza. L'intervallo disponibile è da 50,1% a 99,9%.

Valori mancanti. Consente di specificare le modalità di gestione dei valori mancanti.

Capitolo
43

Strumenti

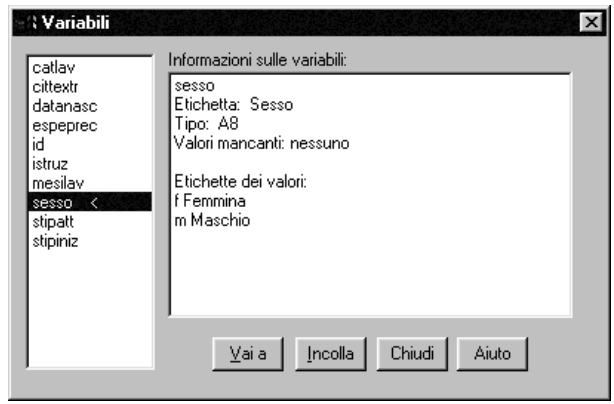
In questo capitolo vengono descritte le funzioni disponibili nel menu Strumenti e la possibilità di riordinare gli elenchi di variabili di destinazione utilizzando i menu di sistema di Windows.

Informazioni sulle variabili

Nella finestra di dialogo Variabili vengono visualizzate informazioni sulla definizione della variabile correntemente selezionata. Tali informazioni includono:

- Formato
- Etichetta della variabile
- Valori mancanti definiti dall'utente
- Etichette dei valori

Figura 43-1
Finestra di dialogo Variabili



Vai a. Passa alla variabile selezionata nella finestra dell'Editor di dati.

Incolla. Incolla le variabili selezionate all'interno della finestra di sintassi designata nel punto in cui è posizionato il cursore.

Per modificare le definizioni delle variabili, utilizzare la Visualizzazione variabili nell'Editor dei dati.

Per ottenere informazioni sulle variabili

- Dai menu, scegliere:
  Strumenti
  Variabili...
- Selezionare la variabile per cui si desidera visualizzare informazioni sulla relativa definizione.

Commenti sul file di dati

È possibile includere commenti descrittivi in un file di dati. Per i file di dati in formato SPSS, questi commenti vengono salvati con il file di dati.

Per aggiungere, modificare, eliminare o visualizzare i commenti sul file di dati

- Dai menu, scegliere:
  Strumenti
  Commenti sul file dati...
- Per visualizzare i commenti nel Viewer, selezionare Visualizza i commenti nell'output.

Le dimensioni dei commenti sono illimitate, tuttavia le singole righe dei commenti non possono superare 80 byte, corrispondenti a 80 caratteri nelle lingue a un byte. Alle righe con dimensioni superiori verrà applicato il ritorno a capo dopo l'ottantesimo carattere. I commenti vengono visualizzati con lo stesso tipo di carattere dell'output testuale, in modo da rifletterne la modalità di visualizzazione nel Viewer.

Un indicatore di data (la data corrente tra parentesi) viene automaticamente aggiunto alla fine della lista di commenti ogni volta che si aggiungono o si modificano i commenti. In questo modo, tuttavia, se si modifica un commento esistente o si inserisce un nuovo commento tra commenti esistenti, è possibile che le date associate ai commenti risultino ambigue.

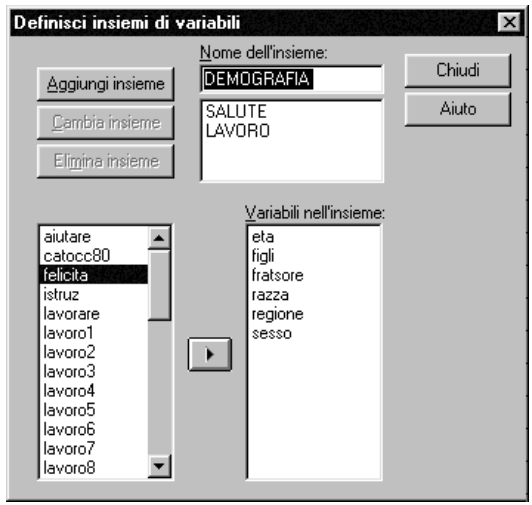
Insiemi di variabili

È possibile limitare il numero di variabili che compaiono negli elenchi di variabili sorgente delle finestre di dialogo mediante la definizione e l'utilizzo di insiemi di variabili. Ciò risulta particolarmente utile per i file di dati che contengono un numero elevato di variabili. L'utilizzo di piccoli insiemi di variabili semplifica la ricerca e la selezione delle variabili per l'analisi e può inoltre migliorare le prestazioni. Se il file di dati include numerose variabili e finestre di dialogo che richiedono tempi di apertura elevati, la limitazione degli elenchi sorgente delle finestre di dialogo a sottoinsiemi più ridotti di variabili permette di diminuire la quantità di tempo richiesta per l'apertura delle finestre di dialogo.

Definisci insiemi di variabili

In questa finestra di dialogo è possibile creare sottoinsiemi di variabili che verranno visualizzati negli elenchi sorgente delle finestre di dialogo.

Figura 43-2
Finestra di dialogo Definisci insiemi di variabili



Nome dell'insieme. I nomi di insiemi possono avere una lunghezza massima di 12 caratteri. È possibile utilizzare qualsiasi carattere, inclusi gli spazi. Per i nomi di insiemi non è attiva la distinzione tra maiuscole e minuscole.

Variabili nell'insieme. In un insieme è possibile inserire qualsiasi combinazione di variabili numeriche, variabili stringa brevi e variabili stringa lunghe. L'ordine delle variabili nell'insieme non influenza in alcun modo l'ordine di visualizzazione delle variabili negli elenchi sorgente delle finestre di dialogo. Una variabile può appartenere a più insiemi di variabili.

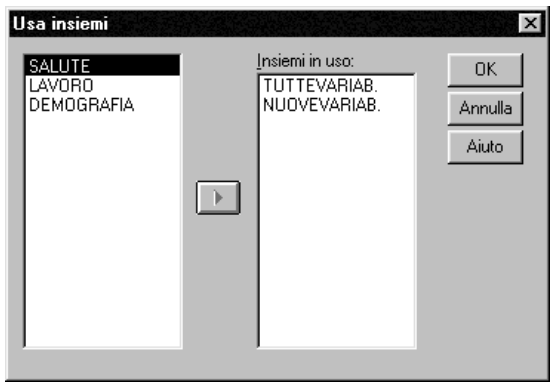
Per definire insiemi di variabili

- Dai menu, scegliere:
  Strumenti
  Definisci insiemi...
- Selezionare le variabili che si desidera includere nell'insieme.
- Immettere un nome per l'insieme (fino a 12 caratteri).
- Fare clic su Aggiungi insieme.

Usa insiemi

Questa finestra di dialogo consente di limitare le variabili visualizzate negli elenchi sorgente delle finestre di dialogo agli insiemi definiti selezionati.

Figura 43-3
Finestra di dialogo Usa insiemi



Insiemi in uso. Visualizza gli insiemi utilizzati per produrre gli elenchi di variabili sorgente nelle finestre di dialogo. Le variabili vengono visualizzate negli elenchi sorgente in ordine alfabetico o di file. L'ordine degli insiemi e l'ordine delle variabili

all'interno di un insieme non influenza in alcun modo l'ordine delle variabili nell'elenco sorgente. Per default, vengono utilizzati due insiemi di sistema:

TUTTEVARIAB. Questo insieme contiene tutte le variabili nel file di dati, incluse le nuove variabili create durante una sessione.

NUOVEVARIAB. Questo insieme contiene solo le nuove variabili create durante la sessione.

È possibile rimuovere questi insiemi dall'elenco e selezionarne altri, ma nell'elenco deve essere disponibile almeno un insieme. Se l'insieme *TUTTEVARIAB* non viene rimosso dall'elenco Insiemi in uso, ogni altro insieme inserito sarà ignorato.

Per limitare gli elenchi sorgente delle finestre di dialogo agli insiemi di variabili definiti

- Dai menu, scegliere:

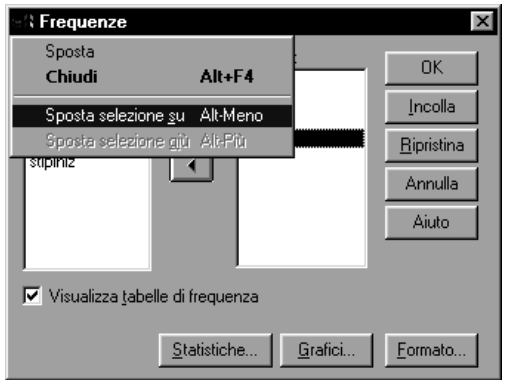
  Strumenti
  Usa insiemi...

- Selezionare gli insiemi di variabili definiti che includono le variabili che si desidera visualizzare negli elenchi sorgente delle finestre di dialogo.

Riordinamento degli elenchi di variabili di destinazione

Le variabili compaiono negli elenchi di destinazione della finestra di dialogo nello stesso ordine in cui vengono selezionate negli elenchi sorgente. Per modificare l'ordine delle variabili in un elenco di destinazione senza dover deselezionare e quindi riselezionare tutte le variabili nel nuovo ordine, è possibile spostare le variabili verso l'alto e verso il basso nell'elenco di destinazione utilizzando il menu di sistema posizionato nell'angolo superiore sinistro della finestra di dialogo (a cui si accede facendo clic sul lato sinistro della barra del titolo della finestra).

Figura 43-4
Menu di sistema di Windows con riordinamento degli elenchi di destinazione



Sposta selezione su. Sposta la variabile o le variabili selezionate di una posizione verso l'alto nell'elenco di destinazione.

Sposta selezione giù. Sposta la variabile o le variabili selezionate di una posizione verso il basso nell'elenco di destinazione.

È possibile spostare simultaneamente più variabili adiacenti (raggruppate insieme). Non è possibile spostare gruppi di variabili non adiacenti.

Capitolo
44

Opzioni

Nella finestra di dialogo Opzioni è possibile impostare numerosi elementi, ad esempio:

- Il giornale di sessione, che consente di registrare tutti i comandi eseguiti in ciascuna sessione.
- L'ordine di visualizzazione delle variabili negli elenchi sorgente delle finestre di dialogo.
- Gli elementi visualizzati e nascosti nei risultati del nuovo output.
- TableLook per le nuove tabelle pivot e ChartLook per i nuovi grafici interattivi
- I formati di valuta personalizzati
- I file e le funzioni di script automatici che consentono di personalizzare l'output

Per modificare le impostazioni delle opzioni

▶ Dai menu, scegliere:

Modifica
  Opzioni...

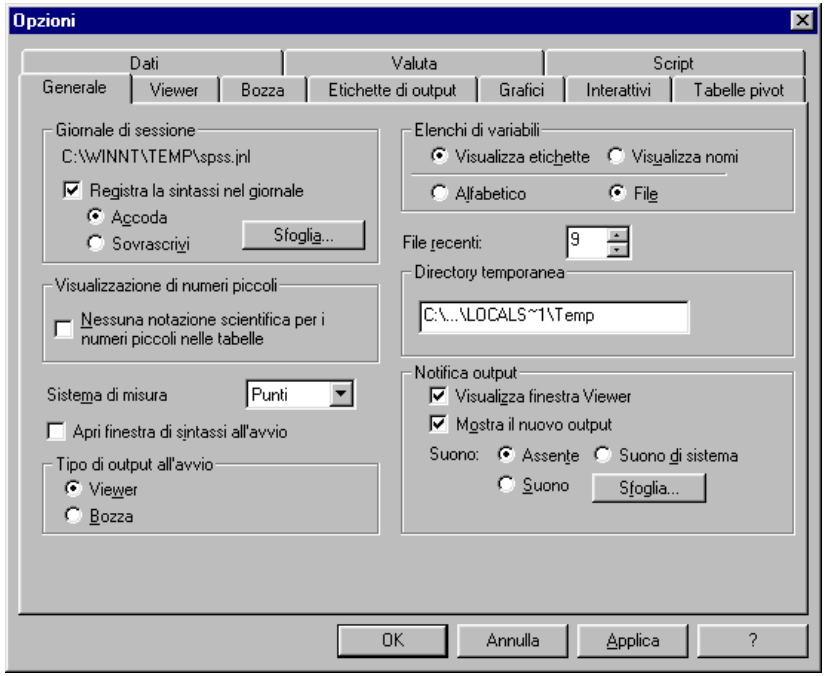
▶ Fare clic sulle schede corrispondenti alle impostazioni che si desidera modificare.

▶ Modificare le impostazioni.

▶ Fare clic su OK o su Applica.

Opzioni generali

Figura 44-1
Finestra di dialogo Opzioni, scheda Generale



Elenchi di variabili. Consente di impostare la visualizzazione delle variabili nelle caselle di riepilogo delle finestre di dialogo. È possibile visualizzare i nomi o le etichette delle variabili. I nomi o le etichette delle variabili possono essere visualizzati in ordine alfabetico o per file, ovvero nell'ordine in cui compaiono nel file di dati e in cui vengono visualizzati nella finestra dell'Editor dei dati. L'ordine di visualizzazione viene applicato solo agli elenchi di variabili sorgente. Gli elenchi di variabili di destinazione riflettono sempre l'ordine di selezione delle variabili.

Giornale di sessione. Il file giornale che registra tutti i comandi eseguiti in ciascuna sessione. Vengono registrati anche i comandi inseriti ed eseguiti nelle finestre di sintassi e quelli generati dalle impostazioni delle finestre di dialogo. È possibile modificare il file giornale e riutilizzare i comandi in altre sessioni di SPSS. È possibile attivare e disattivare la registrazione del giornale, accodare o sovrascrivere il file giornale e selezionare il nome di file e la posizione del giornale. È possibile copiare

la sintassi dei comandi dal file giornale e salvarla in un file di sintassi per utilizzarla con l'Utilità SPSS Production automatica.

Directory temporanea. Consente di determinare la posizione dei file temporanei creati nel corso di una sessione. In modalità distribuita, disponibile con la versione server, ciò non influisce sulla posizione dei file di dati temporanei. In modalità distribuita, la posizione dei file di dati temporanei è controllata dalla variabile di ambiente *SPSSTMPDIR*, che può essere impostata esclusivamente nel computer in cui è in esecuzione la versione server del programma. Se si desidera modificare la posizione della directory temporanea, contattare l'amministratore del sistema.

File recenti. Consente di controllare il numero degli ultimi file utilizzati che compaiono nel menu File.

Apri finestra di sintassi all'avvio. Le finestre di sintassi sono finestre di file di testo in cui è possibile immettere, modificare o eseguire i comandi. Se si utilizza spesso la sintassi dei comandi, selezionare questa opzione per aprire automaticamente una finestra di sintassi all'avvio di ogni sessione. Ciò risulta utile soprattutto per gli utenti esperti che preferiscono utilizzare la sintassi dei comandi anziché le finestre di dialogo. (Non disponibile nella versione per studenti).

Nessuna notazione scientifica per i numeri piccoli nelle tabelle. Consente di sopprimere la visualizzazione della notazione scientifica per valori decimali piccoli nell'output. I valori decimali molto piccoli verranno visualizzati come 0 (oppure 0,000).

Tipo di Viewer all'avvio. Consente di impostare il tipo di Viewer utilizzato e il formato dell'output. Il Viewer produce tabelle pivot e grafici interattivi. La Bozza converte le tabelle pivot in output testuale e i grafici in metafile.

Sistema di misurazione. Il sistema di misura utilizzato (punti, pollici o centimetri) per specificare attributi quali, ad esempio, i margini delle celle delle tabelle pivot, la larghezza delle celle e lo spazio tra le tabelle per la stampa.

Lingua. Consente di controllare la lingua utilizzata nell'output. Non è disponibile per l'output testuale semplice, i grafici interattivi o le mappe, è disponibile con il modulo aggiuntivo Mappe. L'elenco di lingue disponibili dipende dai file di lingua correntemente installati.

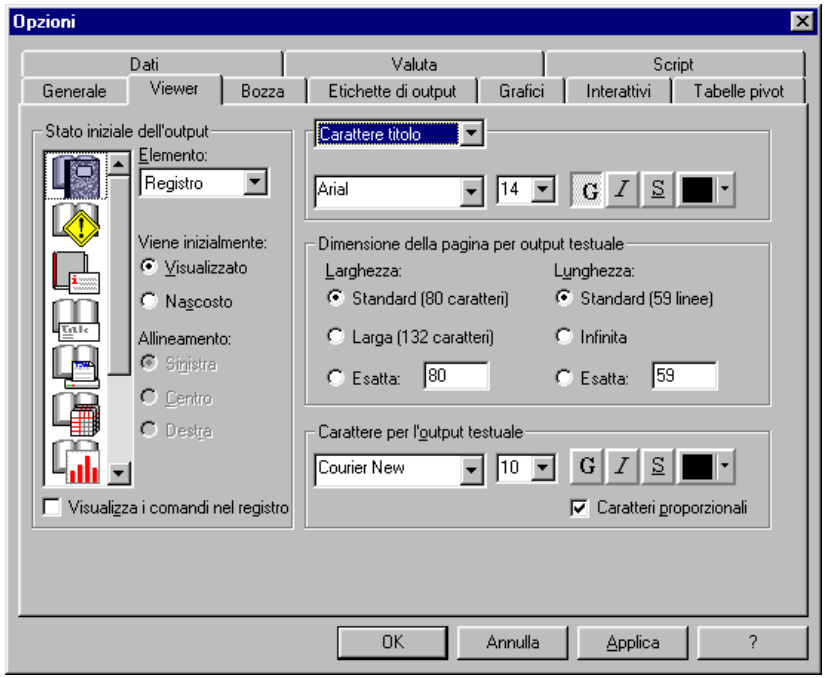
Nota: gli script personalizzati che si basano su stringhe di testo specifiche per la lingua nell'output potrebbero non essere eseguiti correttamente quando viene modificata la lingua dell'output. Per ulteriori informazioni, vedere "Opzioni: Script" a pag. 648.

Notifica. Consente di impostare le modalità di notifica automatica del completamento dell'esecuzione di una procedura e della disponibilità dei risultati nel Viewer.

Opzioni Viewer

Le opzioni di visualizzazione dell'output del Viewer influiscono solo sul nuovo output prodotto dopo la modifica delle impostazioni. Le modifiche apportate a tali impostazioni non influiscono sull'output già visualizzato nel Viewer.

Figura 44-2
Finestra di dialogo Opzioni, scheda Viewer



Stato iniziale dell'output. Consente di impostare gli elementi che verranno visualizzati o nascosti automaticamente ad ogni esecuzione di una procedura e le modalità di allineamento iniziale di tali elementi. È possibile impostare la visualizzazione dei seguenti elementi: registro, avvisi, note, titoli, tabelle pivot, grafici e output testuale (output non visualizzato nelle tabelle pivot). È inoltre possibile attivare o disattivare la visualizzazione dei comandi del registro. È possibile copiare la sintassi dei

comandi dal registro e salvarla in un file di sintassi per utilizzarla con l'Utilità SPSS Production automatica.

Nota: Nel Viewer tutti gli elementi dell'output vengono allineati a sinistra. Le impostazioni di giustificazione vengono applicate solo all'output stampato. Gli elementi centrati e allineati a destra sono contrassegnati da un piccolo simbolo situato sopra e a sinistra di ciascun elemento.

Carattere titolo. Consente di impostare lo stile, la dimensione e il colore del carattere per i titoli del nuovo output.

Carattere titolo pagina. Consente di impostare lo stile, la dimensione e il colore del carattere per i titoli della nuova pagina e per i titoli della pagina generati dalla sintassi dei comandi `TITLE` e `SUBTITLE` o creati tramite l'opzione Nuovo titolo pagina nel menu Inserisci.

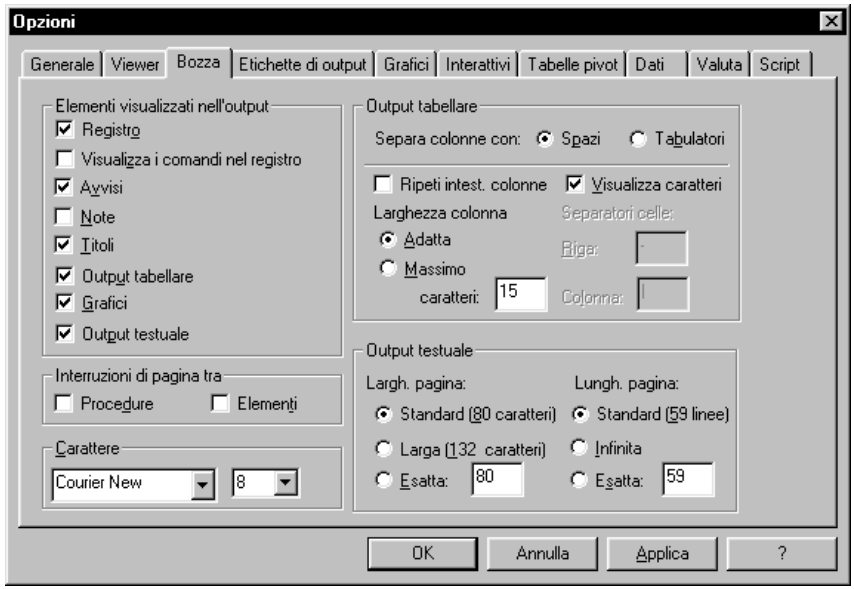
Dimensione della pagina per output testuale. Per l'output testuale, consente di impostare la larghezza (espressa in numero di caratteri) e la lunghezza (espressa in numero di righe) della pagina. Per alcune procedure alcune statistiche vengono visualizzate solo in formato allargato.

Carattere per l'output testuale. Carattere utilizzato per l'output testuale. Per l'output testuale è necessario utilizzare un carattere proporzionale (a spaziatura fissa). Se si seleziona un carattere non proporzionale, l'output non risulterà allineato correttamente nelle tabelle. Il carattere dell'output di testo viene sempre usato dalla procedura guidata per visualizzare i contenuti del file. Tuttavia, il testo non verrà allineato se il carattere ha una spaziatura fissa.

Opzioni Output bozza

Le opzioni di visualizzazione dell'output di Bozza influiscono solo sul nuovo output prodotto dopo che le impostazioni sono state modificate. Le modifiche di tali impostazioni non influiscono sull'output già visualizzato in Bozza.

Figura 44-3
Finestra di dialogo Opzioni, scheda Bozza



Elementi visualizzati nell'output. Consente di impostare gli elementi che verranno visualizzati automaticamente ad ogni esecuzione di una procedura. È possibile impostare la visualizzazione dei seguenti elementi: registro, avvisi, note, titoli, output tabellare (tabelle pivot convertite in output testuale), grafici e output testuale (output delimitato da spazi).È inoltre possibile attivare o disattivare la visualizzazione dei comandi del registro. È possibile copiare la sintassi dei comandi dal registro e salvarla in un file di sintassi per utilizzarla con l'Utilità SPSS Production automatica.

Interruzioni di pagina tra. Consente di inserire le interruzioni di pagina tra gli output di diverse procedure e/o tra i singoli elementi di output.

Carattere. Carattere utilizzato per il nuovo output. Sono disponibili solo i caratteri a spaziatura fissa (proporzionali) in quanto l'output testuale delimitato da spazi non viene allineato correttamente se si utilizza un carattere proporzionale.

Output tabellare. Consente di controllare le impostazioni dell'output delle tabelle pivot convertito in output testuale tabellare. Le specifiche relative alla larghezza e al separatore di colonna sono disponibili solo se si seleziona Spazi come separatore di colonna. Per l'output delle tabelle pivot delimitate da spazi, per impostazione

predefinita il ritorno a capo viene eliminato e per ciascuna colonna viene impostata la larghezza dell'etichetta o del valore più lungo in essa contenuto. Per limitare la larghezza delle colonne e impostare il ritorno a capo automatico per le etichette più lunghe, specificare il numero di caratteri corrispondente alla larghezza desiderata.

Nota: un output separato da tabulatori non verrà tuttavia allineato in modo corretto nella Bozza. Questo formato è utile per copiare e incollare i risultati nelle applicazioni di elaborazione testi, nelle quali è possibile utilizzare qualsiasi tipo di carattere (non solo i caratteri a spaziatura fissa) e impostare le tabulazioni in modo da ottenere un corretto allineamento dell'output.

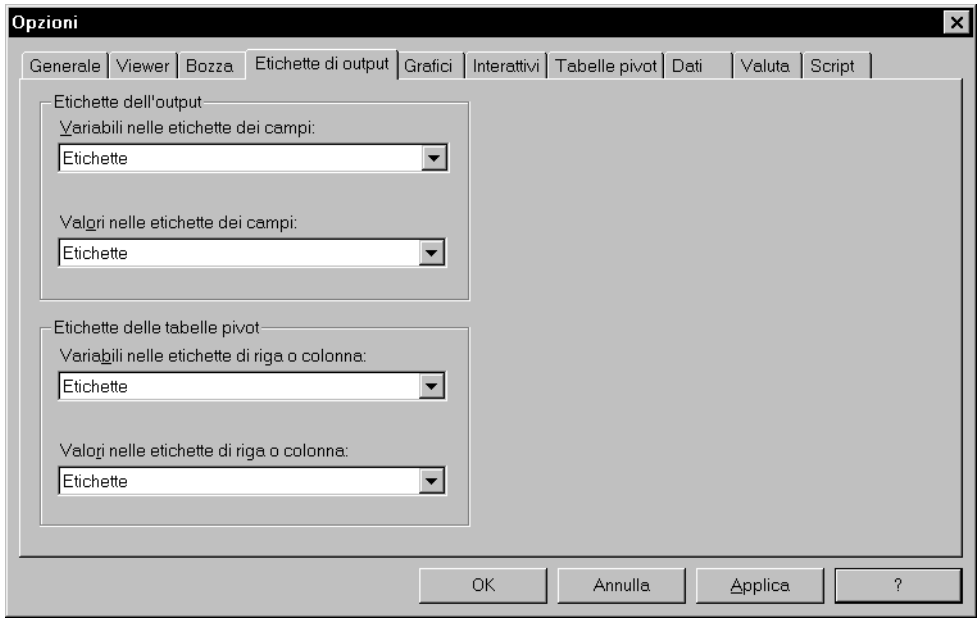
Output testuale. Per l'output testuale diverso dall'output delle tabelle pivot convertite, consente di impostare la larghezza della pagina (espressa in numero di caratteri) e la lunghezza della pagina (espressa in numero di righe). Per alcune procedure alcune statistiche vengono visualizzate solo in formato allargato.

Opzioni: Etichette di output

La scheda Etichette di output consente di impostare la visualizzazione delle informazioni sulle variabili e sui valori dei dati nella vista riassuntiva e nelle tabelle pivot. È possibile visualizzare i nomi delle variabili, le etichette e i valori effettivi delle variabili definite, le etichette dei valori definite o una combinazione.

Le etichette dei valori e di variabili descrittive (Visualizzazione variabili nell'Editor dei dati, colonne *Etichetta* e *Valori*) spesso agevolano l'interpretazione dei risultati. Le etichette molto lunghe, tuttavia, talvolta possono creare dei problemi.

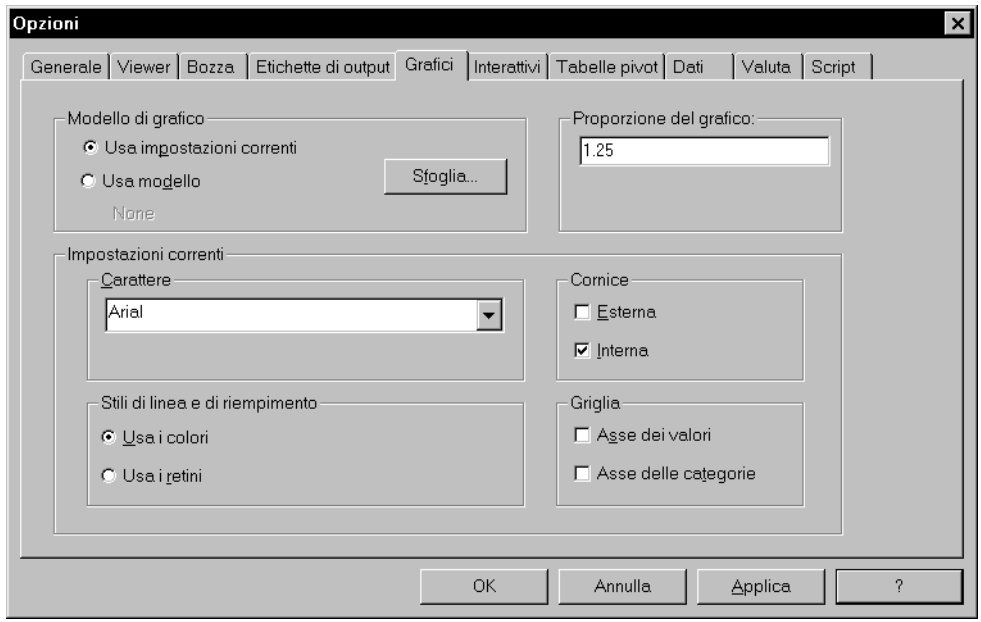
Figura 44-4
Finestra di dialogo Opzioni, scheda Etichette di output



Le opzioni delle etichette di output vengono applicate solo al nuovo output prodotto in seguito alla modifica delle impostazioni. Le modifiche apportate a tali impostazioni non influiscono sull'output già visualizzato nel Viewer. Queste impostazioni vengono applicate solo all'output delle tabelle pivot. L'output testuale rimarrà invariato.

Opzioni: grafici

Figura 44-5
Finestra di dialogo Opzioni, scheda Grafici



Modello di grafico. Per i nuovi grafici è possibile utilizzare le impostazioni selezionate in questo gruppo oppure le impostazioni di un file di modello di grafico. Fare clic su Sfoglia per selezionare un modello di grafico. Per creare un modello di grafico, è necessario creare un file con gli attributi desiderati e quindi salvarlo come modello (menu File, Salva modello).

Proporzione del grafico. Il rapporto tra altezza e larghezza della cornice esterna dei nuovi grafici. È possibile specificare un rapporto larghezza/altezza compreso tra 0,1 e 10,0. Con valori inferiori a 1, si ottengono grafici con un'altezza maggiore della larghezza. Specificando valori maggiori di 1 si otterranno grafici più larghi che alti. Specificando 1 si otterrà un grafico quadrato. Dopo aver creato un grafico non è possibile modificarne le proporzioni.

Esegui JVM all'avvio. Per utilizzare le funzioni relative ai grafici, è necessario disporre di Java Virtual Machine (JVM). Per impostazione predefinita, JVM viene sempre eseguito all'avvio di una sessione di SPSS. Deselezionando questa opzione,

è possibile eseguire SPSS più velocemente, ma è possibile che si verifichino dei leggieri ritardi quando si utilizza JVM per la prima volta per il Generatore di grafici o la creazione di un grafico nella sessione.

Carattere. Il carattere utilizzato per tutto il testo dei nuovi grafici.

Preferenza ciclo stile. L'assegnazione iniziale di colori e/o retini per i nuovi grafici. Usa solo i colori consente di utilizzare solo i colori per differenziare gli elementi del grafico senza usare i retini. Usa i retini consente di utilizzare solo gli stili di linea, i simboli o i retini per differenziare gli elementi del grafico senza usare i colori.

Cornice. Consente di impostare la visualizzazione di cornici interne ed esterne nei nuovi grafici.

Griglia. Consente di impostare la visualizzazione della griglia dell'asse dei valori e delle categorie nei nuovi grafici.

Cicli stile. Consente di personalizzare i colori, gli stili della linea, i simboli e i retini nei nuovi grafici. È possibile modificare l'ordine dei colori e dei retini utilizzati al momento della creazione di un nuovo grafico.

Nota: tali impostazioni non influiscono sui grafici interattivi (menu Grafici, sottomenu Interattivi).

Colori degli elementi dei dati

È possibile specificare l'ordine in cui si desidera che vengano utilizzati i colori per gli elementi dei dati, ad esempio le barre e i simboli, nel nuovo grafico. I colori vengono utilizzati ogni volta che si seleziona un'opzione che includa il *colore* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico a barre raggruppato con due gruppi e si seleziona Usa i colori nella finestra di dialogo Grafici: Opzioni principale, i primi due colori nell'elenco Grafici raggruppati verranno utilizzati come colori delle barre nel nuovo grafico.

Per modificare l'ordine di utilizzo dei colori

- Selezionare Grafici semplici e quindi selezionare un colore utilizzato per i grafici senza categorie.

- Selezionare Grafici raggruppati per modificare il ciclo dei colori per i grafici con le categorie. Per modificare il colore di una categoria, selezionare una categoria e quindi un colore dalla tavolozza.

  Se lo si desidera, è possibile:

  - Inserire una nuova categoria al di sopra di quella selezionata.
  - Spostare una categoria selezionata.
  - Eliminare una categoria selezionata.
  - Ripristinare la sequenza predefinita.
  - Modificare un colore selezionandolo e quindi scegliendo Modifica.

Linee degli elementi dei dati

È possibile specificare l'ordine in cui devono essere utilizzati gli stili per i dati lineari nel grafico. Gli stili delle linee vengono utilizzati ogni volta che nel grafico sono presenti dati lineari e che si seleziona un'opzione che includa i *retini* nel gruppo Preferenza ciclo stile dnella finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico lineare con due gruppi e si seleziona Usa solo i retini nella finestra di dialogo Grafici: Opzioni principale, i primi due stili nell'elenco Grafici raggruppati verranno utilizzati come retini nel nuovo grafico.

Per modificare l'ordine di utilizzo degli stili delle linee

- Selezionare Grafici semplici e quindi selezionare uno stile di linea utilizzato per i grafici senza categorie.
- Selezionare Grafici raggruppati per modificare il ciclo dei retini per i grafici lineari con le categorie. Per modificare lo stile della linea di una categoria, selezionare una categoria e quindi uno stile dalla tavolozza.

  Se lo si desidera, è possibile:

  - Inserire una nuova categoria al di sopra di quella selezionata.
  - Spostare una categoria selezionata.
  - Eliminare una categoria selezionata.
  - Ripristinare la sequenza predefinita.

Simboli degli elementi dei dati

È possibile specificare l'ordine in cui devono essere utilizzati i simboli per gli elementi dei dati nel grafico. Gli stili dei simboli vengono utilizzati ogni volta che nel grafico sono presenti simboli e che si seleziona un'opzione che includa i *retini* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico a dispersione con due gruppi e si seleziona Usa solo i retini nella finestra di dialogo Grafici: Opzioni principale, i primi due simboli nell'elenco Grafici raggruppati verranno utilizzati come simboli nel nuovo grafico.

Per modificare l'ordine di utilizzo degli stili dei simboli

- Selezionare Grafici semplici e quindi selezionare un simbolo per i grafici senza categorie.
- Selezionare Grafici raggruppati per modificare il ciclo dei retini per i grafici con le categorie. Per modificare il simbolo di una categoria, selezionare una categoria e quindi un simbolo dalla tavolozza.

Se lo si desidera, è possibile:

- Inserire una nuova categoria al di sopra di quella selezionata.
- Spostare una categoria selezionata.
- Eliminare una categoria selezionata.
- Ripristinare la sequenza predefinita.

Riempimenti degli elementi dei dati

È possibile specificare l'ordine in cui devono essere utilizzati gli stili di riempimento per le aree e le barre nel grafico. Gli stili di riempimento vengono utilizzati ogni volta che nel grafico sono presenti barre o aree e che si seleziona un'opzione che includa i *retini* nel gruppo Preferenza ciclo stile della finestra di dialogo Grafici: Opzioni principale.

Se, ad esempio, si crea un grafico a barre raggruppato con due gruppi e si seleziona Usa solo i retini nella finestra di dialogo Grafici: Opzioni principale, i primi due stili nell'elenco Grafici raggruppati verranno utilizzati come retini di riempimento nel nuovo grafico.

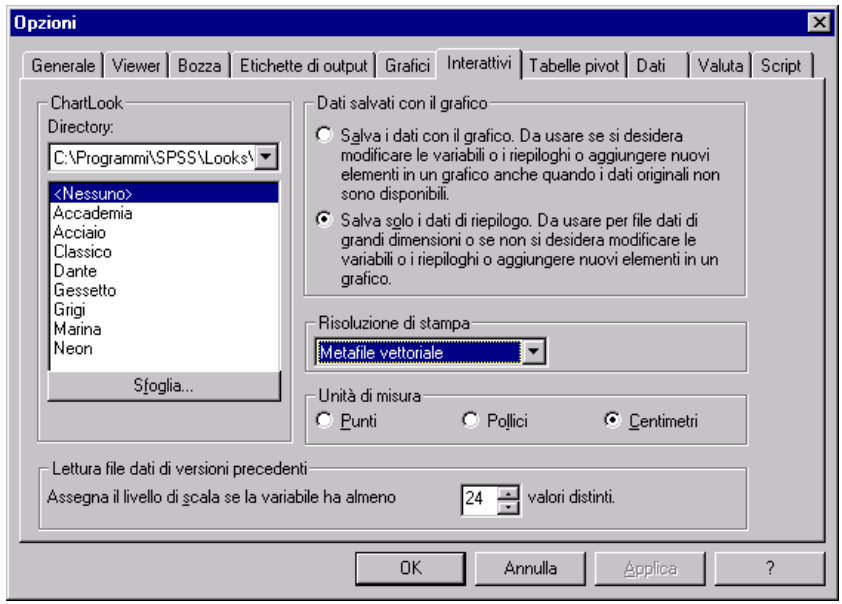
Per modificare l'ordine di utilizzo degli stili di riempimento

- Selezionare Grafici semplici e quindi selezionare un retino utilizzato per i grafici senza categorie.
- Selezionare Grafici raggruppati per modificare il ciclo dei retini per i grafici con le categorie. Per modificare il retino di una categoria, selezionare una categoria e quindi un retino dalla tavolozza.

  Se lo si desidera, è possibile:
  - Inserire una nuova categoria al di sopra di quella selezionata.
  - Spostare una categoria selezionata.
  - Eliminare una categoria selezionata.
  - Ripristinare la sequenza predefinita.

Opzioni dei grafici interattivi

Figura 44-6
Finestra di dialogo Opzioni, scheda Interattivi



Per i grafici interattivi (menu Grafici, sottomenu Interattivi), sono disponibili le seguenti opzioni:

ChartLook. Selezionare un ChartLook dall'elenco dei file e quindi fare clic su OK o su Applica. Per impostazione predefinita, nell'elenco vengono visualizzati i ChartLook salvati nella directory *Looks* della directory in cui è installato il programma. È possibile utilizzare uno dei ChartLook disponibili nel programma oppure crearne uno personalizzato nell'Editor dei grafici interattivi (menu Formato, ChartLook in un grafico attivato).

- **Directory.** Consente di selezionare una directory dei ChartLook. Utilizzare Sfoglia per aggiungere directory all'elenco.
- **Sfoglia.** Consente di selezionare un ChartLook da un'altra directory.

Dati salvati con il grafico. Consente di controllare le informazioni salvate con i grafici interattivi quando i grafici non sono più allegati al file dati di origine (per esempio, se si apre un file del Viewer salvato in una sessione precedente). Salvando i dati con il grafico, è possibile eseguire la maggior parte delle funzioni interattive disponibili per i grafici allegati al file dati di origine (ad eccezione dell'aggiunta delle variabili non incluse nel grafico originale). Questa procedura tuttavia potrebbe aumentare in modo significativo le dimensioni dei file Viewer, in modo particolare nel caso di file dati di grandi dimensioni.

Risoluzione di stampa. Consente di impostare la risoluzione di stampa dei grafici interattivi. Nella maggior parte dei casi, Metafile vettoriale consente di stampare in modo più rapido e di ottenere risultati migliori. Per le bitmap, la stampa di grafici a bassa risoluzione risulterà più veloce, mentre i grafici ad alta risoluzione saranno di migliore qualità.

Unità di misura. Il sistema di misura utilizzato (punti, pollici o centimetri) per specificare attributi quali le dimensioni dell'area dati di un grafico.

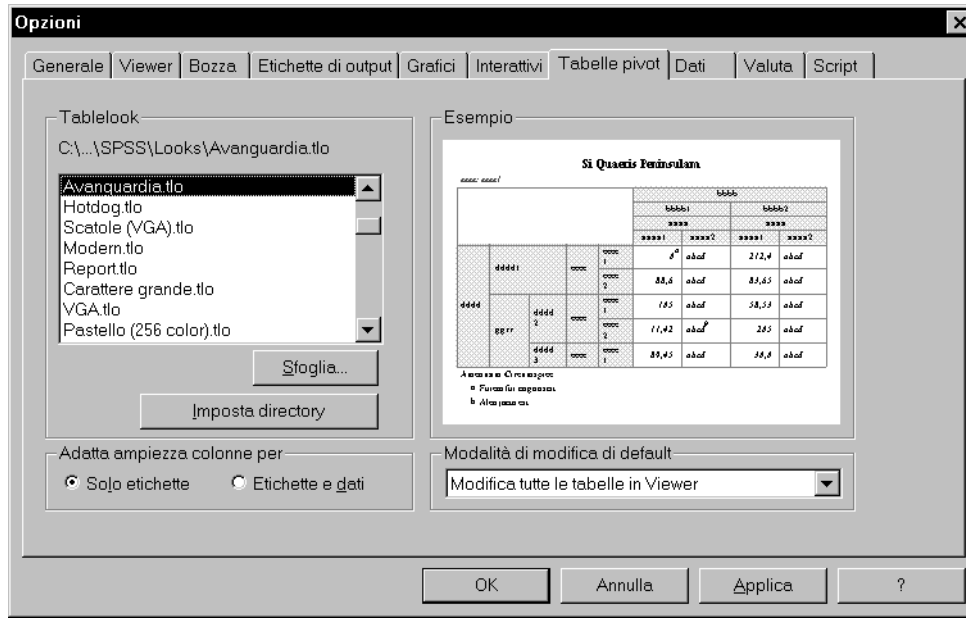
Lettura file dati di versioni precedenti. Per i file di dati creati nelle versioni precedenti di SPSS, i dati letti dai file esterni e le nuove variabili create in una sessione, è possibile specificare il numero minimo di valori dati per la variabile numerica usata per classificare la variabile come scala o nominale. Le variabili con un numero di valori distinti inferiore a quello specificato vengono classificate come nominali.

Nota: queste impostazioni si applicano solo ai grafici interattivi (sottomenu Interattivi del menu Grafici), fatta eccezione per il livello di misurazione.

Opzioni tabella pivot

La scheda Tabella pivot consente di impostare il TableLook predefinito utilizzato per l'output delle nuove tabelle pivot. Utilizzando i TableLook è possibile impostare numerosi attributi delle tabelle pivot, ad esempio la visualizzazione e la larghezza della griglia, lo stile, la dimensione e il colore del carattere e i colori di sfondo.

Figura 44-7
Finestra di dialogo Opzioni, scheda Tabelle pivot



TableLook. Selezionare un TableLook dall'elenco dei file e quindi fare clic su OK o su Applica. Per impostazione predefinita, nell'elenco vengono visualizzati i TableLook salvati nella directory *Looks* della directory in cui è installato il programma. È possibile utilizzare uno dei TableLook disponibili in SPSS oppure crearne uno personalizzato nell'Editor di tabelle pivot (menu Formato, TableLook).

- **Sfoglia.** Consente di selezionare un TableLook da un'altra directory.
- **Imposta directory.** Consente di modificare la directory predefinita dei TableLook.

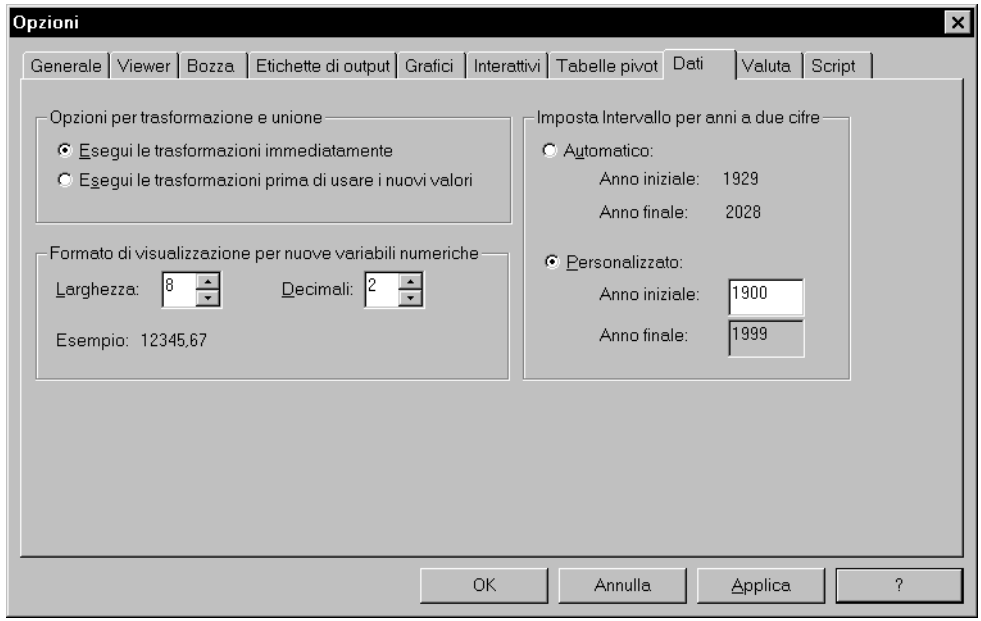
Adatta ampiezza colonne per. Consente di modificare l'adattamento automatico della larghezza delle colonne delle tabelle pivot.

- **Solo etichette.** Consente di adattare la larghezza delle colonne a quella delle relative etichette. In questo modo è possibile ottenere tabelle più compatte, ma i valori di larghezza maggiore di quella delle etichette non verranno visualizzati (gli asterischi indicano i valori troppo grandi per poter essere visualizzati).
- **Etichette e dati.** Consente di adattare la larghezza delle colonne in base alle etichette delle colonne o ai valori, a seconda di quale sia il maggiore. In questo modo si otterranno tabelle più grandi in cui, tuttavia, verranno visualizzati tutti i valori.

Modalità di modifica di default. Consente di impostare l'attivazione delle tabelle pivot nella finestra del Viewer o in una finestra separata. Per impostazione predefinita, facendo doppio clic su una tabella pivot si attiva la tabella nella finestra del Viewer. È possibile attivare le tabelle pivot in una finestra separata oppure selezionare un'impostazione di dimensioni per aprire tabelle pivot più piccole nella finestra del Viewer e tabelle pivot più grandi in una finestra separata.

Opzioni: Dati

Figura 44-8
Finestra di dialogo Opzioni, scheda Dati



Opzioni per trasformazione e unione. Ogni volta che viene eseguito un comando, viene letto il file dati. Per alcune trasformazioni di dati (ad esempio Calcola e Ricodifica) e di file (ad esempio Aggiungi variabili e Aggiungi casi) non è necessario il passaggio separato dei dati e l'esecuzione di questi comandi può essere rimandata finché non vengono letti i dati per l'esecuzione di un altro comando, ad esempio una procedura statistica. Per file dati di grandi dimensioni, selezionare Esegui le trasformazioni prima di usare i nuovi valori per rimandare l'esecuzione e risparmiare tempo di elaborazione.

Formato di visualizzazione per nuove variabili numeriche. Consente di modificare la larghezza di visualizzazione predefinita e il numero di decimali delle nuove variabili numeriche. Per le nuove variabili stringa non è previsto un formato di visualizzazione predefinito. Se un valore è troppo grande per il formato di visualizzazione specificato, verranno innanzitutto arrotondate le cifre decimali e quindi i valori verranno convertiti in notazione scientifica. I formati di visualizzazione non producono alcun effetto sui

valori interni. Ad esempio, il valore 123456,78 può essere arrotondato a 123457 per la visualizzazione, ma nei calcoli verrà utilizzato il valore originale non arrotondato.

Imposta intervallo per anni a due cifre. Consente di definire l'intervallo di anni per le variabili in formato data immesse e/o visualizzate con un anno a due cifre (per esempio, 28/10/86, 29-OTT-87). L'impostazione automatica dell'intervallo è basata sull'anno corrente, con inizio 69 anni prima e con termine 30 anni dopo l'anno corrente (aggiungendo l'anno corrente si ottiene un intervallo totale di 100 anni). Per un intervallo personalizzato, l'anno finale viene determinato automaticamente in base al valore immesso come anno iniziale.

Generatore di numeri casuale. Sono disponibili due diversi generatori di numeri casuali:

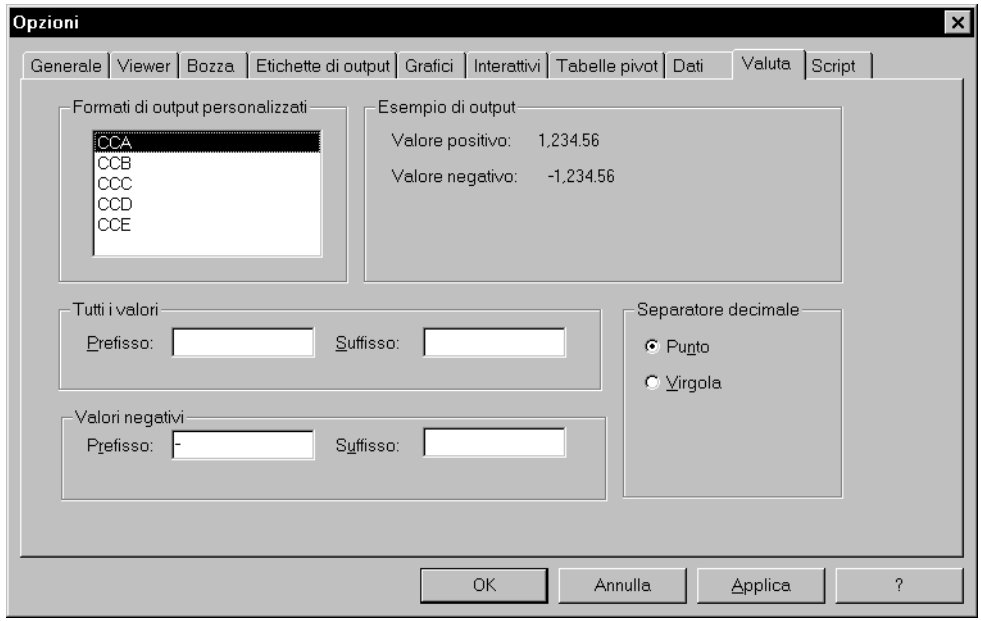
- **Compatibile con SPSS 12.** Il generatore di numeri casuali usato in SPSS 12 e versioni precedenti. Usare questo generatore se è necessario riprodurre i risultati casuali ottenuti da versioni precedenti di SPSS.
- **Mersenne Twister.** Un nuovo generatore di numeri casuali più affidabile per scopi di simulazione. Usare questo generatore se non serve riprodurre i risultati casuali ottenuti da versioni precedenti di SPSS.

Opzioni: Valuta

È possibile creare al massimo cinque formati di visualizzazione della valuta personalizzati che possono includere caratteri speciali di prefisso o suffisso e un particolare trattamento dei valori negativi.

I nomi dei cinque formati di valuta personalizzati sono CCA, CCB, CCC, CCD e CCE. Non è possibile modificare i nomi dei formati o aggiungerne altri. Per modificare un formato di valuta personalizzato, selezionarne il nome nell'elenco di origine e apportare le modifiche desiderate.

Figura 44-9
Finestra di dialogo Opzioni, scheda Valuta



I prefissi, i suffissi e i separatori decimali definiti per i formati di valuta personalizzati vengono utilizzati solo ai fini della visualizzazione. Non è possibile inserire valori nell'Editor dei dati utilizzando i caratteri di valuta personalizzati.

Per creare formati di valuta personalizzati

- ▶ Fare clic sulla scheda Valuta.
- ▶ Selezionare uno dei formati valuta dall'elenco (CCA, CCB, CCC, CCD e CCE).
- ▶ Specificare il prefisso, il suffisso e il separatore decimale.
- ▶ Fare clic su OK o su Applica.

Opzioni: Script

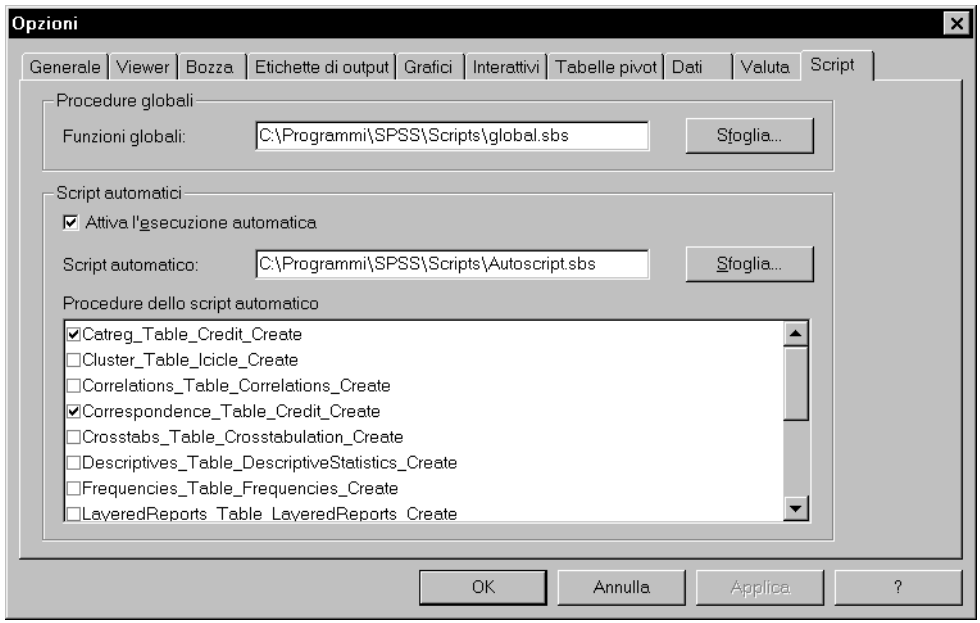
Utilizzare la scheda Script per specificare il file delle procedure globali e il file degli script automatici e quindi selezionare le procedure dello script automatico che si desidera utilizzare. È possibile utilizzare gli script per eseguire automaticamente numerose funzioni, ad esempio la personalizzazione delle tabelle pivot.

Procedure globali. Un file delle procedure globali è una libreria di procedure e funzioni di script che vengono chiamate dai file di script, inclusi i file di script automatici.

Nota: il file delle procedure globali incluso nel programma viene selezionato per impostazione predefinita. Molti script disponibili in SPSS utilizzano le funzioni e le procedure di questo file delle procedure globali e non funzionano se si specifica un file diverso.

Script automatici. Uno script automatico è un insieme di procedure dello script che vengono eseguite automaticamente ogni volta che si eseguono procedure in grado di creare particolari tipi di oggetti di output.

Figura 44-10
Finestra di dialogo Opzioni, scheda Script



Vengono visualizzate tutte le procedure dello script automatico corrente e pertanto è possibile attivare e disattivare singole procedure.

Per specificare i file degli script automatici e delle procedure globali

- Fare clic sulla scheda Script.
- Selezionare Attiva l'esecuzione automatica.
- Selezionare le procedure dello script automatico che si desidera attivare.

  È inoltre possibile specificare un file degli script automatici o delle procedure globali differente.

Capitolo
45

Personalizzazione dei menu e delle barre degli strumenti

Editor dei menu

È possibile utilizzare l'Editor dei menu per personalizzare i menu. Nell'Editor dei menu è possibile:

- Aggiungere comandi di menu per l'esecuzione di script personalizzati.
- Aggiungere comandi di menu per l'esecuzione di file della sintassi dei comandi.
- Aggiungere comandi di menu per l'avvio di altre applicazioni e l'invio automatico di dati ad altre applicazioni.

È possibile inviare dati ad altre applicazioni nei seguenti formati: SPSS, Excel 4.0, Lotus 1-2-3 versione 3, SYLK, tab-delimitato e dBASE IV.

Per aggiungere comandi ai menu

▶ Dai menu, scegliere:
Strumenti
Editor dei menu...

▶ Nella finestra di dialogo Editor dei menu fare doppio clic sul menu a cui si desidera aggiungere un comando.

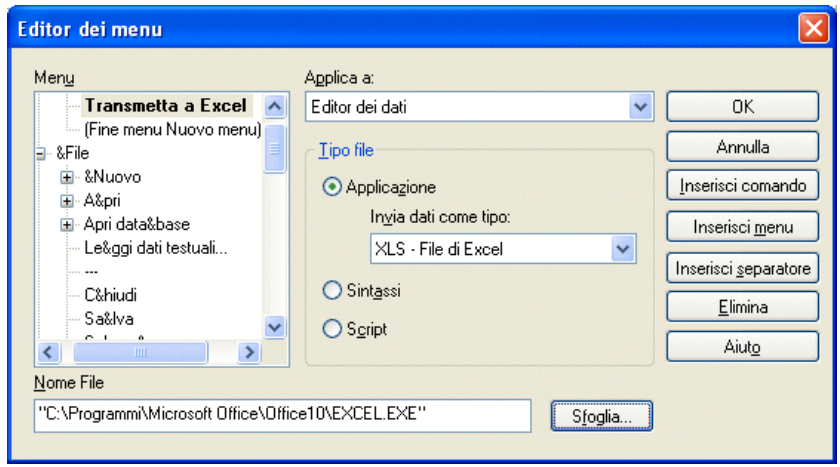
▶ Selezionare il comando del menu che si desidera preceda il nuovo comando.

▶ Fare clic su Inserisci comando per inserire un nuovo comando nel menu.

▶ Selezionare il tipo di file per il nuovo comando (file script, file della sintassi dei comandi o applicazione esterna).

- Fare clic su Sfoglia per selezionare un file da associare al nuovo comando di menu.

Figura 45-1
Finestra di dialogo Editor dei menu



È inoltre possibile aggiungere menu completamente nuovi e separatori tra i comandi dei menu.

Nonché inviare automaticamente il contenuto dell'Editor dei dati a un'altra applicazione selezionandola dai menu.

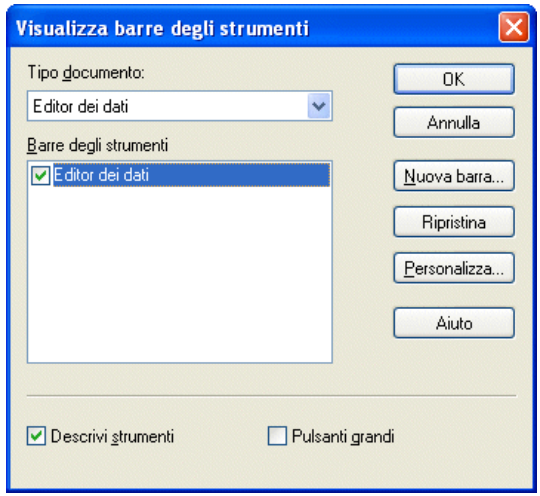
Personalizzazione delle barre degli strumenti

È possibile personalizzare le barre degli strumenti e crearne di nuove. Le barre degli strumenti possono contenere tutti gli strumenti disponibili, inclusi quelli per le azioni dei menu. Possono inoltre esservi inclusi strumenti personalizzati che permettono di avviare altre applicazioni, eseguire file di sintassi o eseguire file di script.

Visualizza barre degli strumenti

La finestra Visualizza barre degli strumenti consente di visualizzare o nascondere le barre degli strumenti, personalizzarle oppure crearne di nuove. Le barre degli strumenti possono contenere tutti gli strumenti disponibili, inclusi quelli per le azioni dei menu. Possono inoltre esservi inclusi strumenti personalizzati che permettono di avviare altre applicazioni, eseguire file di sintassi o eseguire file di script.

Figura 45-2
Finestra di dialogo Visualizza barre degli strumenti



Per personalizzare le barre degli strumenti

- Dai menu, scegliere:
  Visualizza
  Barre degli strumenti...
- Selezionare la barra degli strumenti che si desidera personalizzare e quindi fare clic su Personalizza oppure su Nuova barra per creare una nuova barra degli strumenti.
- Per le nuove barre degli strumenti, specificare il nome, selezionare le finestre in cui si desidera inserire la nuova barra e quindi fare clic su Personalizza.
- Selezionare una voce dell'elenco Categorie per visualizzare gli strumenti disponibili nella categoria desiderata.
- Trascinare gli strumenti desiderati sulla barra degli strumenti visualizzata nella finestra di dialogo.

- Per rimuovere uno strumento, trascinarlo all'esterno della barra degli strumenti visualizzata nella finestra di dialogo.

Per creare uno strumento personalizzato per l'apertura di file, l'esecuzione di file di sintassi o di script:

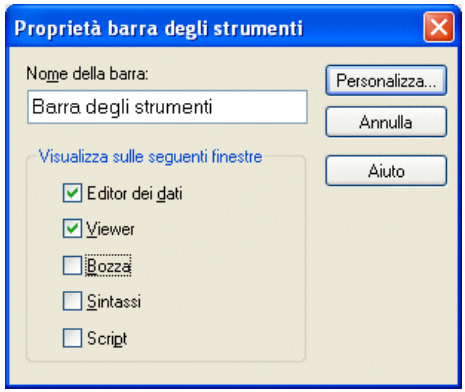
- Fare clic su Nuovo strumento nella finestra di dialogo Personalizza barra degli strumenti.
- Digitare un'etichetta descrittiva per lo strumento.
- Selezionare il comando che si desidera assegnare allo strumento (apertura di file, esecuzione di file di sintassi o di script).
- Fare clic su Sfoglia per selezionare un file o un'applicazione da associare allo strumento.

I nuovi strumenti verranno visualizzati nella categoria Personalizzati, che contiene anche i comandi di menu definiti dall'utente.

Proprietà barra degli strumenti

Nella finestra di dialogo Proprietà barra degli strumenti è possibile selezionare i tipi di finestra in cui si desidera visualizzare la barra degli strumenti selezionata. In questa finestra di dialogo è inoltre possibile creare i nomi delle nuove barre degli strumenti.

Figura 45-3
Finestra di dialogo Proprietà barra degli strumenti



Per impostare le proprietà delle barre degli strumenti

- Dai menu, scegliere:

  Visualizza
  Barre degli strumenti...

- Per le barre degli strumenti esistenti, fare clic su Personalizza e quindi su Proprietà nella finestra di dialogo Personalizza barra degli strumenti.

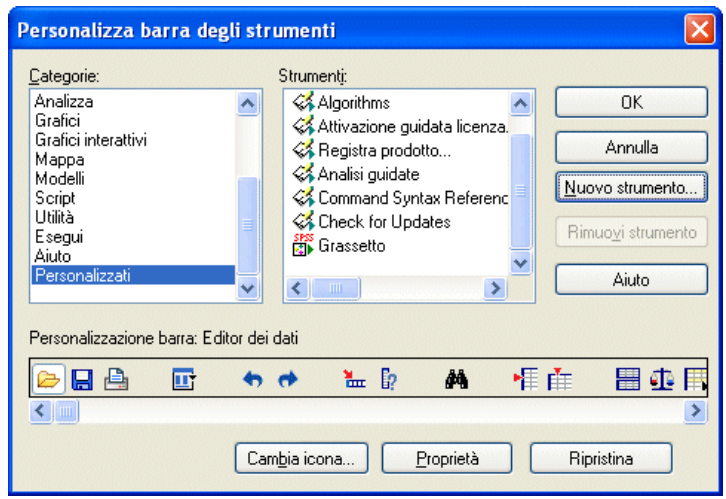
- Per le nuove barre degli strumenti, fare clic su Nuovo strumento.

- Selezionare i tipi di finestra in cui si desidera visualizzare la barra degli strumenti. Per le nuove barre degli strumenti, specificare anche il nome.

Personalizza barra degli strumenti

Nella finestra di dialogo Personalizza barra degli strumenti è possibile creare nuove barre degli strumenti e personalizzare quelle esistenti. Le barre degli strumenti possono contenere tutti gli strumenti disponibili, inclusi quelli per le azioni dei menu. Possono inoltre esservi inclusi strumenti personalizzati che permettono di avviare altre applicazioni, eseguire file di sintassi o eseguire file di script.

Figura 45-4
Finestra di dialogo Personalizza barra degli strumenti



Crea nuovo strumento

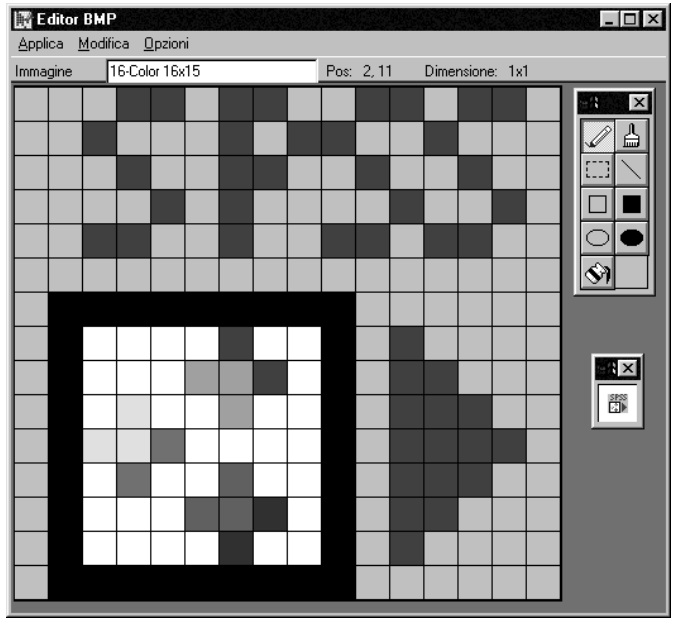
Nella finestra di dialogo Crea nuovo strumento è possibile creare strumenti personalizzati per l'avvio di altre applicazioni, eseguire file di sintassi e file di script.

Figura 45-5
Finestra di dialogo Crea nuovo strumento

Editor di bitmap

Utilizzare l'Editor BMP per creare icone personalizzate per i pulsanti della barra degli strumenti. Ciò risulta particolarmente utile per gli strumenti personalizzati creati per l'esecuzione di script, sintassi e altre applicazioni.

Figura 45-6
Editor BMP



Per modificare le bitmap della barra degli strumenti

- Dai menu, scegliere:
  Visualizza
  Barre degli strumenti...

- Selezionare la barra degli strumenti che si desidera personalizzare e quindi fare clic su Personalizza.

- Nella barra degli strumenti di esempio fare clic sullo strumento di cui si desidera modificare l'icona.

- Fare clic su Cambia icona.

- Utilizzare la casella degli strumenti e la tavolozza dei colori per modificare l'icona o per crearne una nuova.

Capitolo
46

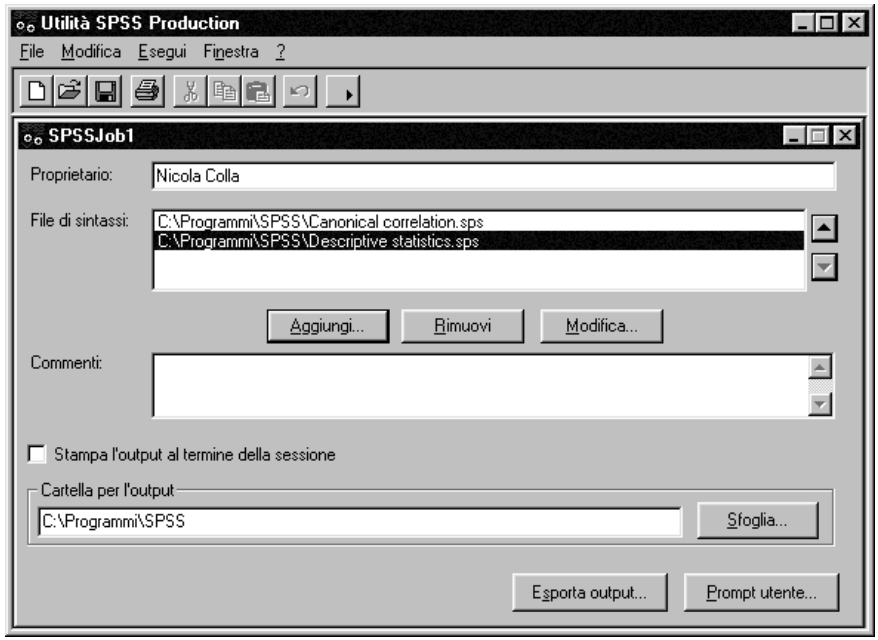
Utilità SPSS Production

L'utilità SPSS Production consente di eseguire SPSS in modo automatico. Il programma viene eseguito in modo automatico e termina dopo l'esecuzione dell'ultimo comando. In tale modo, si possono svolgere contemporaneamente altre attività. L'utilità Production è utile nel caso di esecuzione ripetuta degli stessi insiemi di analisi che richiedono tempi di elaborazione piuttosto lunghi, per esempio i report settimanali.

L'utilità SPSS Production utilizza i file della sintassi dei comandi per indicare al programma le operazioni da eseguire. Un file della sintassi dei comandi è un semplice file di testo che contiene la sintassi dei comandi. Per creare il file è possibile utilizzare qualsiasi editor di testo. È inoltre possibile generare la sintassi dei comandi incollando le selezioni della finestra di dialogo in una finestra di sintassi o modificando il file giornale.

Dopo aver creato i file di sintassi e averli inclusi in una sessione dell'utilità SPSS Production, è possibile visualizzarli e modificarli nell'utilità stessa.

Figura 46-1
utilità SPSS Production



Formato di input della sintassi. Controlla la forma delle regole della sintassi utilizzate per il lavoro:

- **Interattive.** Ciascun comando deve terminare con un punto. I punti possono essere inseriti in qualsiasi punto del comando. Sebbene i comandi possano continuare su più righe, l'uso di punto come ultimo carattere non vuoto viene interpretato come la fine di un comando. Le righe di continuazione ed i nuovi comandi possono iniziare in qualsiasi punto della nuova riga. Queste regole "interattive" si applicano quando si selezionano e si eseguono comandi in una finestra di sintassi.
- **Batch.** Ciascun comando deve iniziare all'inizio di una riga (non devono essere inseriti spazi vuoti prima dell'inizio del comando), mentre le linee di continuazione devono essere rientrate di almeno uno spazio. Per rientrare i nuovi comandi, è possibile utilizzare il segno più ("+"), la barra o un punto come primo carattere all'inizio della riga, quindi rientrare il comando vero e proprio. Il punto alla fine del comando è facoltativo. Questa impostazione è compatibile con le regole della sintassi dei file dei comandi inclusi con il comando `INCLUDE`.

Errore di sintassi: comportamento. Controlla la modalità con cui vengono gestite le condizioni di errore della sessione:

- **Continua.** Gli errori della sessione non interrompono automaticamente l'elaborazione dei comandi. I comandi dei file della sessione vengono considerati parte del normale flusso dei comandi, quindi l'elaborazione dei comandi viene continuata normalmente.
- **Interrompi.** L'elaborazione dei comandi si interrompe appena si verifica il primo errore nella sessione. Ciò è compatibile con il comportamento dei file dei comandi inclusi con il comando `INCLUDE`.

Risultati della sessione dell'utilità SPSS Production. Ad ogni esecuzione dell'utilità SPSS Production viene creato un file di output con lo stesso nome della sessione dell'utilità e l'estensione *spo*. Ad esempio, il file di una sessione dell'utilità SPSS Production denominato *prodjob.spp* consente di creare il file di output *prodjob.spo*. Il file di output è un documento del Viewer di SPSS.

Tipo di output. L'output del Viewer produce tabelle pivot e grafici interattivi ad alta risoluzione. L'output della Bozza produce output di testo e immagini metafile dei grafici. Nella Bozza è possibile modificare l'output di testo, ma non i grafici.

Per utilizzare l'utilità SPSS Production

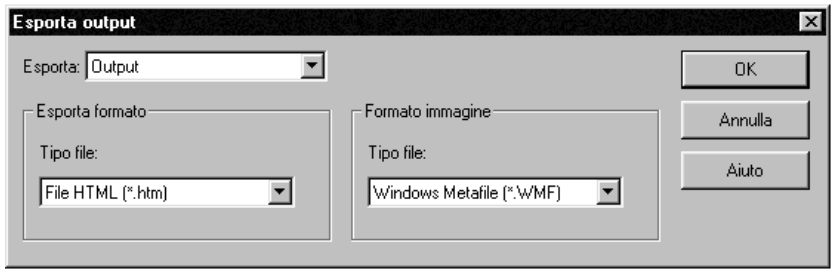
- ▶ Creare un file di sintassi dei comandi.
- ▶ Avviare l'utilità SPSS Production, disponibile nel menu Start.
- ▶ Specificare i file di sintassi che si desidera utilizzare per la sessione dell'utilità SPSS Production. Fare clic su Aggiungi per selezionarli.
- ▶ Salvare il file della sessione dell'utilità SPSS Production.
- ▶ Eseguire il file della sessione dell'utilità SPSS Production. Fare clic sul pulsante Esegui sulla barra degli strumenti oppure dai menu, scegliere:

  Esegui
  Sessione Utilità SPSS Production

Opzioni di esportazione

L'opzione Opzioni (Esportazione) consente di salvare tabelle pivot e output testuale in formato HTML, di testo, Word/RTF ed Excel, nonché di salvare i grafici in diversi formati comuni utilizzati da altre applicazioni.

Figura 46-2
Finestra di dialogo Opzioni (Esportazione)



Esporta

In questo elenco a discesa viene specificato l'elemento da esportare.

Output. Consente di esportare qualsiasi combinazione di tabelle pivot, output testuale e grafici.

- Per i formati HTML e di testo, i grafici vengono esportati nel formato di esportazione correntemente selezionato. Per il formato HTML, i grafici vengono incorporati come indirizzi e possono essere esportati in un formato adatto all'inserimento in documenti HTML. Per il formato di testo, nel file di testo viene inserita una linea per ogni grafico in cui viene indicato il nome di file del grafico esportato.
- Per il formato Word/RTF, i grafici vengono esportati in formato metafile di Windows e incorporati nel documento Word.
- I grafici non vengono inclusi in documenti Excel.

Output (senza grafici). Consente di esportare tabelle pivot e output testuale. I grafici nel Viewer vengono ignorati.

Solo grafici. Consente di esportare solo grafici. Per i documenti HTML e in formato testo, i formati di esportazione includono: metafile avanzati (EMF), metafile di Windows (WMF), bitmap di Windows (BMP), PostScript encapsulated (EPS), JPEG,

TIFF, PNG e PICT di Macintosh. Per i documenti Word/RTF, i grafici vengono sempre esportati in formato metafile di Windows.

Esporta formato

Per i documenti di output, le opzioni disponibili sono HTML, testo, Word/RTF ed Excel. Per i formati HTML e testo, i grafici vengono invece esportati nel formato di grafico correntemente selezionato. Per l'opzione Solo grafici, selezionare un formato di esportazione dall'elenco a discesa. Per i documenti di output, le tabelle e il testo vengono esportati nel seguente modo:

- **File HTML (*.htm).** Le tabelle pivot vengono esportate come tabelle HTML. L'output testuale è esportato come HTML preformattato.
- **File di testo (*.txt).** È possibile esportare le tabelle pivot in formato delimitato da tabulazioni o da spazi. Tutto l'output testuale viene esportato in formato delimitato da spazi.
- **File Excel (*.xls).** Le righe, le colonne e le celle delle tabelle pivot vengono esportate come righe, colonne e celle di Excel, con tutti gli attributi di formattazione, ad esempio, con i bordi delle celle, gli stili del carattere, il colore di sfondo delle celle e così via. L'output testuale viene esportato con tutti gli attributi del carattere. Ogni riga nell'output testuale è una riga del file di Excel, con tutti i contenuti di una riga raccolti in una singola cella.
- **File Word/RTF (*.doc).** Le tabelle pivot vengono esportate come tabelle Word con tutti gli attributi di formattazione; ad esempio bordi delle celle, stili del carattere, colori dello sfondo e così via. L'output testuale viene esportato come HTML formattato. L'output testuale in SPSS viene sempre visualizzato con caratteri a spaziatura fissa (uno spazio) e viene esportato con gli stessi attributi del carattere. Per il corretto allineamento dell'output testuale delimitato da spazi è necessario utilizzare un carattere a spaziatura fissa (proporzionale).

Formato immagine

Consente di impostare il formato di esportazione dei grafici. È possibile esportare grafici nei seguenti formati: metafile avanzati (EMF), metafile di Windows (WMF), bitmap di Windows (BMP), PostScript encapsulated (EPS), JPEG, TIFF, PNG o PICT di Macintosh.

I nomi dei grafici esportati sono formati dal nome di file della sessione dell'utilità SPSS Production seguito da un numero progressivo, con l'estensione del formato selezionato. Ad esempio, se tramite sessione dell'utilità SPSS Production *prodjob.spp* vengono esportati grafici in formato metafile di Windows, ai grafici vengono assegnati i nomi *prodjob1.wmf, prodjob2.wmf, prodjob3.wmf* e così via.

Opzioni per il testo e per le immagini

Le opzioni per l'esportazione di testo (ad esempio delimitato da tabulazioni o da spazi) e le opzioni per l'esportazione di grafici (ad esempio impostazioni di colore, dimensioni e risoluzione) vengono impostate in SPSS e non possono essere modificate nell'utilità SPSS Production. Scegliere Esporta dal menu File di SPSS per modificare le opzioni di esportazione di testo e grafici.

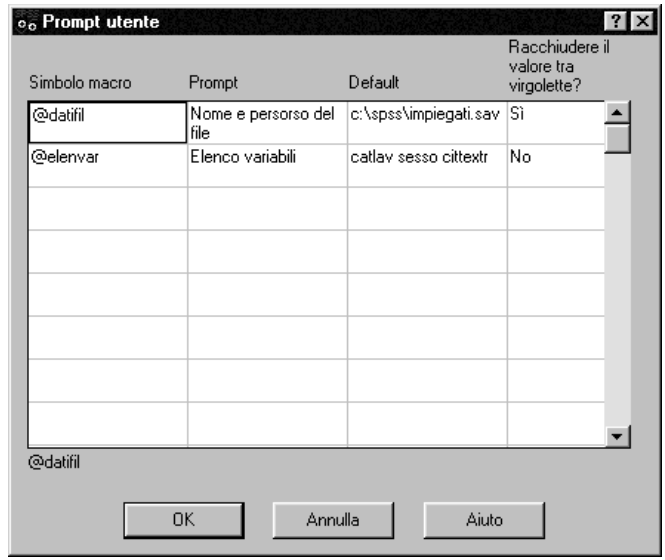
Opzione di esportazione del componente Bozza

L'unica opzione di esportazione disponibile per l'output della Bozza è quella relativa all'output in formato di testo semplice. Non è possibile esportare i grafici per l'output della Bozza.

Prompt utente

I simboli macro definiti in un file della sessione dell'utilità SPSS Production e utilizzati in un file di sintassi consentono di semplificare operazioni come l'esecuzione della stessa analisi per file di dati diversi o l'esecuzione di una serie di comandi per diversi insiemi di variabili. Ad esempio, è possibile definire il simbolo macro *@filedat* in modo che venga richiesto il nome di un file dati ad ogni esecuzione di una sessione dell'utilità SPSS Production in cui venga utilizzata la stringa *@filedat* al posto di un nome di file nel file della sintassi del comando.

Figura 46-3
Finestra di dialogo Prompt utente



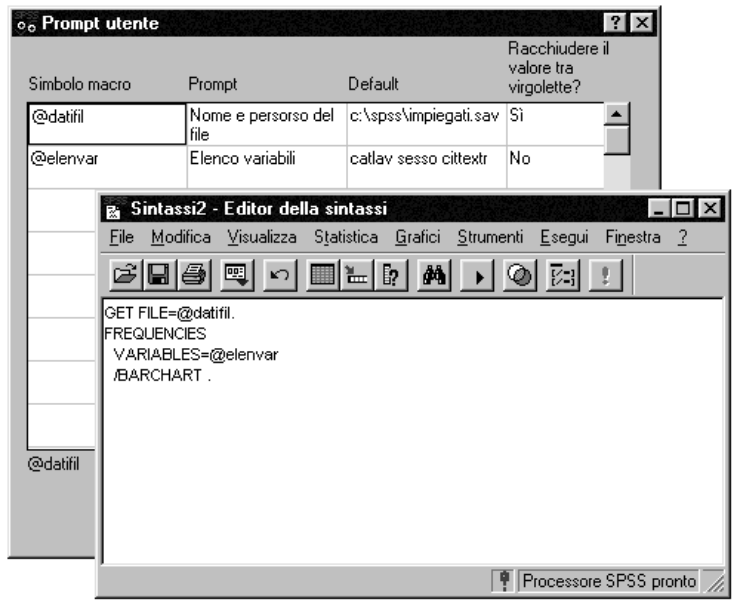
Simbolo macro. Il nome della macro utilizzato nel file di sintassi per richiamare la macro che richiede l'inserimento di informazioni da parte dell'utente. Il simbolo macro deve iniziare con il carattere @. *Nota*: questi simboli di "macro" non sono correlati a quelle che possono essere create con l'utility di SPSS tramite `DEFINE-!ENDDEFINE`.

Prompt. L'etichetta descrittiva visualizzata quando nella sessione dell'utilità SPSS Production vengono richieste informazioni all'utente. Ad esempio, è possibile utilizzare l'espressione "Quale file dati utilizzare?" per identificare un campo che richiede l'indicazione del nome di un file dati.

Default. Il valore fornito per impostazione predefinita dalla sessione dell'utilità SPSS Production se non viene specificato un valore diverso. Questo valore viene visualizzato quando nella sessione dell'utilità SPSS Production viene richiesto l'inserimento di informazioni. È possibile sostituire o modificare il valore in fase di esecuzione.

Racchiudere il valore tra virgolette. Specificare S o Sì se si desidera racchiudere il valore tra virgolette. In caso contrario, lasciare vuoto il campo o specificare N o No. Ad esempio, si consiglia di immettere Sì per specificare un nome di file, poiché i nomi di file devono essere inclusi tra virgolette.

Figura 46-4
Prompt macro in un file di sintassi



Prompt macro dell'utilità SPSS Production

L'utilità SPSS Production richiede di inserire valori ad ogni esecuzione di una sessione dell'utilità contenente simboli macro definiti. È possibile sostituire o modificare i valori predefiniti visualizzati. In SPSS tali valori verranno sostituiti ai simboli macro in tutti i file di sintassi associati alla sessione dell'utilità SPSS Production.

Figura 46-5
Finestra di dialogo Prompt macro dell'utilità SPSS Production

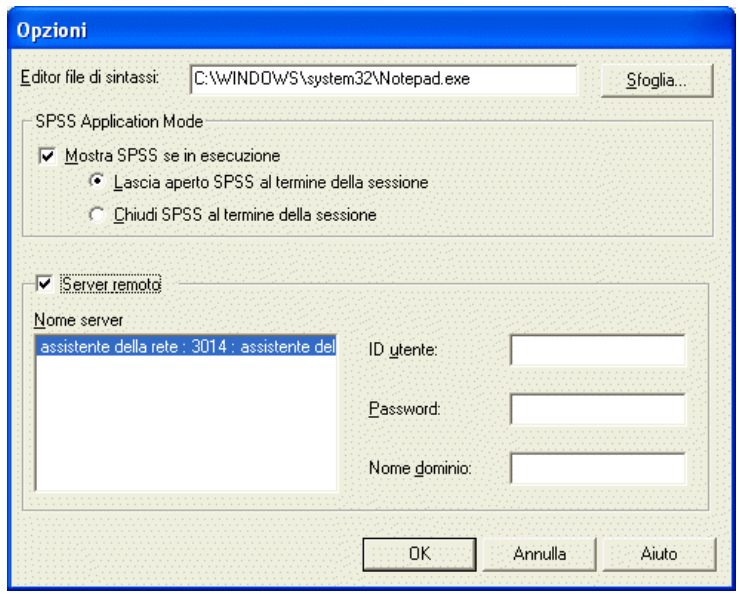
Nota: questi simboli di "macro" non sono correlati a quelle che possono essere create con l'utilità di SPSS tramite `DEFINE-!ENDDEFINE` e l'utilità SPSS Production li richiede anche nel caso in cui nessuno dei file delle sintassi dei comandi del lavoro contenga riferimenti alle "macro" definite.

Opzioni dell'utilità SPSS Production

Le opzioni dell'utilità SPSS Production consentono di:

- Specificare un editor di testo predefinito per i file di sintassi a cui si accede con il pulsante Modifica nella finestra di dialogo principale.
- Eseguire la sessione dell'utilità SPSS Production in background oppure visualizzare SPSS e i risultati generati al progredire della sessione.
- Specificare server remoto, nome di dominio, ID utente e password per l'analisi distribuita (solo nel caso in cui si disponga di un accesso di rete alla versione server di SPSS). Se non si specificano tali impostazioni, vengono utilizzate le impostazioni predefinite della finestra di dialogo Login server di SPSS. È possibile selezionare solo i server remoti definiti in precedenza nella finestra di dialogo Aggiungi server in SPSS (menu File, Cambia server, Aggiungi).

Figura 46-6
Finestra di dialogo Opzioni



Per modificare le opzioni dell'utilità SPSS Production

Dai menu dell'utilità SPSS Production, scegliere:

Modifica
  Opzioni...

Impostazione del formato per le sessioni dell'utilità SPSS Production

Molte impostazioni di SPSS consentono di specificare il formato migliore per le tabelle pivot create durante le sessioni dell'utilità SPSS Production:

TableLook. Modificando e salvando i TableLook (menu Formato di una tabella pivot attivata) è possibile impostare molti attributi delle tabelle pivot. È possibile specificare dimensioni e stili di carattere, colori e bordi. Per fare in modo che le tabelle di grandi dimensioni non vengano divise da un'interruzione di pagina, selezionare Ridimensiona la tabella in larghezza nella scheda Generale della finestra di dialogo Proprietà tabella.

Etichette di output. Le opzioni Etichette di output (menu Modifica, Opzioni, scheda Etichette di output) consentono di impostare la visualizzazione delle informazioni sulle variabili e sui valori nelle tabelle pivot. È possibile visualizzare i nomi delle variabili e/o le etichette delle variabili definite, i valori effettivi e/o le etichette dei valori definite. Le etichette descrittive dei valori e delle variabili semplificano spesso l'interpretazione dei risultati, tuttavia le etichette lunghe potrebbero essere visualizzate in modo non corretto nelle tabelle.

Larghezza colonna. Le opzioni della tabella pivot (menu Modifica, Opzioni, scheda Tabella pivot) consentono di impostare il TableLook predefinito e la regolazione automatica della larghezza delle colonne nelle tabelle pivot.

- **Solo etichette.** Consente di adattare la larghezza delle colonne a quella delle relative etichette. In questo modo è possibile ottenere tabelle più compatte, ma i valori di larghezza maggiore di quella delle etichette non verranno visualizzati (gli asterischi indicano i valori troppo grandi per poter essere visualizzati).
- **Etichette e dati.** Consente di adattare la larghezza delle colonne in base alle etichette delle colonne o ai valori, a seconda di quale sia il maggiore. In questo modo si otterranno tabelle più grandi in cui, tuttavia, verranno visualizzati tutti i valori.

Per TableLook e opzioni, nelle sessioni dell'utilità SPSS Production verranno utilizzate le impostazioni valide in SPSS. È possibile definire le impostazioni desiderate per TableLook e Opzioni in SPSS prima di eseguire la sessione dell'utilità SPSS Production oppure definirle utilizzando i comandi `SET` nella sintassi. Utilizzando i comandi `SET` nei file di sintassi è possibile utilizzare più impostazioni di TableLook e Opzioni nella stessa sessione.

Per creare un TableLook predefinito personalizzato

- ▶ Attivare una tabella pivot (fare doppio clic su un punto qualsiasi della tabella).
- ▶ Dai menu, scegliere:
  Formato
  TableLook...
- ▶ Selezionare un TableLook dall'elenco e quindi fare clic su Modifica.
- ▶ Impostare le proprietà della tabella indicando gli attributi desiderati.
- ▶ Fare clic su Salva o Salva con nome per salvare il TableLook e quindi fare clic su OK.

- Dai menu, scegliere:
  Modifica
  Opzioni...
- Fare clic sulla scheda Tabelle pivot.
- Selezionare il TableLook dall'elenco e quindi fare clic su OK.

Per impostare le opzioni per le sessioni dell'utilità SPSS Production

- Dai menu, scegliere:
  Modifica
  Opzioni...
- Selezionare le opzioni desiderate.
- Fare clic su OK.

  Nella finestra di dialogo Opzioni è possibile impostare il TableLook predefinito, le etichette di output e l'adattamento automatico della larghezza delle colonne. Le impostazioni di Opzioni vengono salvate insieme al programma. Per la sessione dell'utilità SPSS Production, verranno applicate le impostazioni di Opzioni specificate l'ultima volta che è stato eseguito SPSS.

Impostazione del formato delle tabelle pivot con la sintassi

SET TLOOK. Consente di impostare il TableLook predefinito per le nuove tabelle pivot, ad esempio:

```
SET TLOOK = 'c:\prodjobs\mytable.tlo'.
```

SET TVARS. Consente di impostare la visualizzazione dei nomi e delle etichette delle variabili nelle nuove tabelle pivot.

- `SET TVARS = LABELS` consente di visualizzare le etichette delle variabili.
- `SET TVARS = NAMES` consente di visualizzare i nomi delle variabili.
- `SET TVARS = BOTH` consente di visualizzare sia i nomi sia le etichette delle variabili.

SET ONUMBER. Consente di impostare la visualizzazione dei valori o delle etichette dei valori nelle nuove tabelle pivot.

- `SET ONUMBER = LABELS` consente di visualizzare le etichette dei valori.
- `SET ONUMBER = VALUES` consente di visualizzare i valori dei dati.
- `SET ONUMBER = BOTH` consente di visualizzare i valori dei dati e le etichette dei valori.

SET TFIT. Consente di impostare l'adattamento automatico della larghezza delle colonne per le nuove tabelle pivot.

- `SET TFIT = LABELS` consente di adattare la larghezza delle colonne a quella delle relative etichette.
- `SET TFIT = BOTH` consente di adattare la larghezza delle colonne a quella delle relative etichette o del valore dei dati più grande, a seconda di quale sia il maggiore.

Esecuzione di sessioni dell'utilità SPSS Production dalla riga di comando

Le opzioni della riga di comando consentono di programmare l'esecuzione delle sessioni dell'utilità SPSS Production agli orari specificati grazie a utilità di pianificazione come quella disponibile in Microsoft Plus!. Per eseguire le sessioni dell'utilità SPSS Production dalla riga di comando, utilizzare le seguenti opzioni:

-r. Consente di eseguire la sessione dell'utilità SPSS Production. Se la sessione contiene prompt utente, è necessario inserire le informazioni richieste prima di eseguirla.

-s. Consente di eseguire la sessione dell'utilità SPSS Production eliminando tutti i prompt utente e i messaggi. Verranno utilizzati automaticamente i valori predefiniti dei prompt utente.

Analisi distribuita. Se si dispone dell'accesso in rete alla versione server di SPSS, è inoltre possibile utilizzare le seguenti opzioni per eseguire l'utilità SPSS Production in modalità di analisi distribuita:

-x. Nome o indirizzo IP del server remoto.

-n. Numero porta.

-d. Nome del dominio.

-u. ID utente per l'accesso al server remoto.

-p. Password per l'accesso al server remoto.

Se si specifica una qualsiasi delle opzioni della riga di comando per l'analisi distribuita, è necessario specificare anche tutte le altre opzioni (`-x`, `-n`, `-d`, `-u` e `-p`).

È necessario specificare l'intero percorso dell'utilità SPSS Production (*spssprod.exe*) e della sessione dell'utilità, entrambi tra virgolette. Ad esempio:

```
"c:\programmi\spss\spssprod.exe"  "c:\spss\datajobs\prodjob.spp" -s -r
```

Per le opzioni della riga di comando che necessitano di ulteriori specifiche, è necessario aggiungere all'opzione un segno di uguale immediatamente seguito dalle ulteriori specifiche. Se le specifiche contengono degli spazi (ad esempio, un nome di server composto da due parole), racchiudere il valore tra virgolette o apostrofi, come nell'esempio seguente:

```
-x="HAL 9000" -u="secret word"
```

Server predefinito. Se si dispone dell'accesso in rete alla versione server di SPSS, il server predefinito e le relative informazioni (se non specificate nelle opzioni della riga di comando) vengono visualizzate nella finestra di dialogo Login server di SPSS. Se non vi sono impostazioni predefinite visualizzate, la sessione viene eseguita in modalità locale.

Se si desidera eseguire una sessione dell'utilità SPSS Production in modalità locale, ma il computer locale non corrisponde al server predefinito, specificare delle stringhe vuote tra virgolette per tutte le opzioni della riga di comando relative all'analisi distribuita, come nell'esempio seguente:

```
"c:\programmi\spss\spssprod.exe"  "c:\spss\datajobs\prodjob.spp" -x="" -n="" -d="" -u="" -p=""
```

Esecuzione di più sessioni

Se si utilizza un file batch (.bat) o un'utilità simile per eseguire più sessioni, è necessario utilizzare il comando `Start`di Windows con l'opzione `/wait` per controllare l'esecuzione di ciascuna sessione ed evitare che le sessioni successive inizino prima della fine della sessione precedente; come in:

```
cd \program files\spss
start /wait spssprod.exe prodjob1.spp -s
start /wait spssprod.exe prodjob2.spp -s
```

Pubblica sul Web

Pubblica sul Web esporta l'output per la pubblicazione su SmartViewer Web Server. È possibile visualizzare e manipolare sul Web le tabelle e i report pubblicati in SmartViewer in tempo reale con un normale browser.

- Le tabelle pivot sono pubblicate come tabelle dinamiche che possono essere manipolate sul Web per ottenere diverse visualizzazioni dei dati.
- I grafici sono pubblicati come file grafici JPEG o PNG.
- L'output testuale è pubblicato come HTML preformattato. (Per default, la maggior parte dei browser Web utilizza caratteri a dimensioni fisse per i testi preformattati).

Pubblica. Consente di specificare l'output che si desidera pubblicare:

- **Output.** Consente di pubblicare l'intero documento di output, inclusi gli elementi nascosti o compressi.
- **Output (senza note).** Consente di pubblicare tutto eccetto le tabelle delle note create automaticamente per ciascuna procedura.
- **Solo tabelle.** Consente di escludere i grafici. Vengono pubblicate tutte le tabelle pivot e le tabelle di testo.
- **Solo tabelle (senza note).** Consente di escludere i grafici e le tabelle delle note.
- **Solo grafici.** Consente di pubblicare solo i grafici all'interno del documento.
- **Non pubblicare.** Consente di annullare la pubblicazione sul Web. L'esecuzione della sessione dell'utilità salva automaticamente tutte le impostazioni (file (*spp*), di conseguenza i risultati verranno pubblicati a ogni esecuzione della sessione a

meno che non si scelga Non pubblicare. Tale opzione annulla la pubblicazione, pur generando output di altro tipo (file Viewer, file HTML) specificati nella sessione dell'utilità.

Pubblica tabelle come. Consente di controllare le modalità di pubblicazione delle tabelle pivot:

- **Interattive.** Le tabelle sono oggetti dinamici che possono essere manipolati sul Web per ottenere diverse visualizzazioni dei dati.
- **Statiche.** Le tabelle sono statiche e non possono essere manipolate dopo la pubblicazione.

Configura. Consente di aprire la pagina di configurazione della pubblicazione automatica del server Web SmartViewer in una finestra del browser. Ciò è necessario se si apre una nuova sessione dell'utilità per la pubblicazione sul Web.

Per accedere al server Web SmartViewer sono inoltre necessari ID utente e password. Se si apre una nuova sessione dell'utilità per la pubblicazione sul Web, vengono richiesti ID utente e password. Le informazioni vengono memorizzate nella sessione in formato crittografato.

Nota: l'opzione Pubblica sul Web è disponibile solo per i siti con il server Web SmartViewer installato e richiede un plugin per attivare la funzione di pubblicazione. Rivolgersi all'amministratore di sistema o al webmaster per istruzioni sul prelevamento del plugin. Se SmartViewer non è disponibile, usare Esporta output per salvare l'output in formato HTML.

SmartViewer Web Server: Login

La pubblicazione sul server Web SmartViewer richiede un nome utente (ID utente) e una password validi per il server Web SmartViewer.

Rivolgersi all'amministratore di sistema o al webmaster per ulteriori informazioni.

Capitolo
47

Utilità Script di SPSS

Gli script disponibili in SPSS consentono di automatizzare le seguenti operazioni:

- Apertura e salvataggio dei file di dati
- Visualizzazione e gestione delle finestre di dialogo
- Esecuzione di trasformazioni di dati e procedure statistiche usando la sintassi dei comandi
- Esportazione di grafici come file grafici in formati diversi
- Personalizzazione dell'output nel Viewer

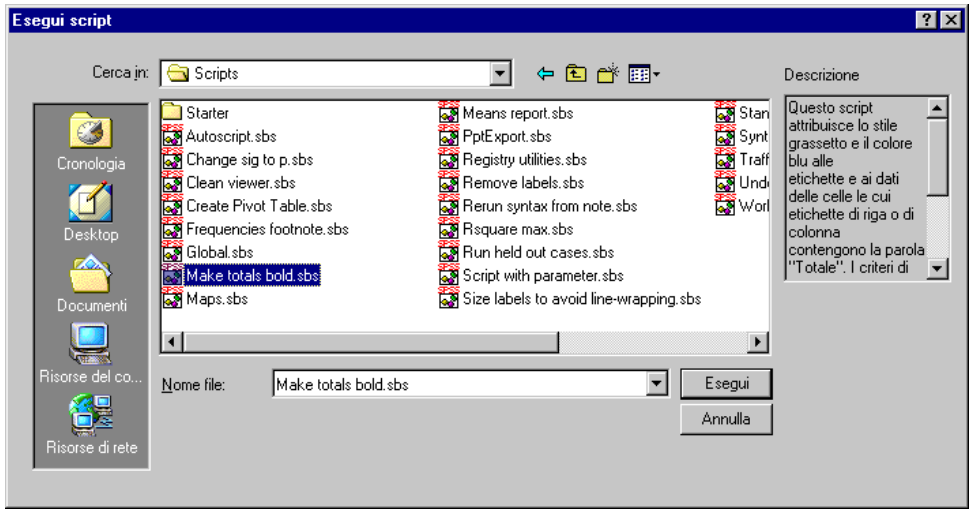
Il software include alcuni script, tra i quali gli script automatici che vengono eseguiti ogni volta che si genera un tipo di output specifico. È possibile usare gli script così come sono oppure personalizzarli in base alle specifiche esigenze. Se si desidera creare nuovi script, è possibile usare uno degli script di avvio disponibili come punto di partenza.

Per eseguire uno script

▶ Dai menu, scegliere:

Strumenti
  Esegui script...

Figura 47-1
Finestra di dialogo Esegui script



- Selezionare la cartella *Script*.
- Selezionare lo script desiderato.

Per ulteriori informazioni, vedere “Personalizzazione dei menu e delle barre degli strumenti” in Capitolo 45 a pag. 651.

Script disponibili in SPSS

Nel programma sono disponibili i seguenti script:

Esegui su casi esclusi. Consente di ripetere un’analisi fattoriale o discriminante usando casi non selezionati in un’analisi precedente. Per eseguire lo script è necessario selezionare una tabella Note generata con un’analisi fattoriale o discriminante eseguita precedentemente.

Cambia sig di p. Consente di cambiare *Sig.* in *p=* nelle etichette di colonna delle tabelle pivot. Per eseguire lo script, è necessario selezionare la tabella.

Pulisci Viewer. Consente di eliminare tutte le tabelle Note da un output. Per eseguire lo script è necessario aprire il documento nella finestra designata del Viewer.

Note frequenze. Consente di inserire le statistiche visualizzate in una tabella Frequenze: Statistiche come note a piè di pagina della tabella di frequenze corrispondente per ciascuna variabile. Per poter eseguire lo script, è necessario che la tabella Frequenze: Statistiche sia selezionata.

Totali in grassetto. Consente di applicare il formato grassetto e il colore blu a righe, colonne o strati di dati contraddistinti dall'etichetta *Totale* in una tabella pivot. Per eseguire lo script, è necessario selezionare la tabella.

Report medie. Consente di estrarre informazioni da una tabella Medie e di scrivere i risultati in diversi file di output ASCII. Per eseguire lo script è necessario selezionare la tabella Medie.

Rimuovi etichette. Consente di eliminare tutte le etichette di riga e colonna dalla tabella pivot selezionata. Per eseguire lo script, è necessario selezionare la tabella.

Esegui sintassi da note. Consente di rieseguire il comando individuato nella tabella Note selezionata usando il file di dati attivo. Se non è aperto alcun file di dati, viene eseguito un tentativo di lettura del file di dati SPSS usato inizialmente. Per eseguire lo script è necessario selezionare la tabella Note.

R quadrato massimo. In una tabella di riepilogo del modello della regressione consente di applicare il formato grassetto e il colore blu alla riga corrispondente al modello che massimizza l'R^{quadrato} corretto. Per eseguire lo script è necessario selezionare la tabella Riepilogo del modello.

Per ulteriori informazioni, vedere "Opzioni" in Capitolo 44 a pag. 629.

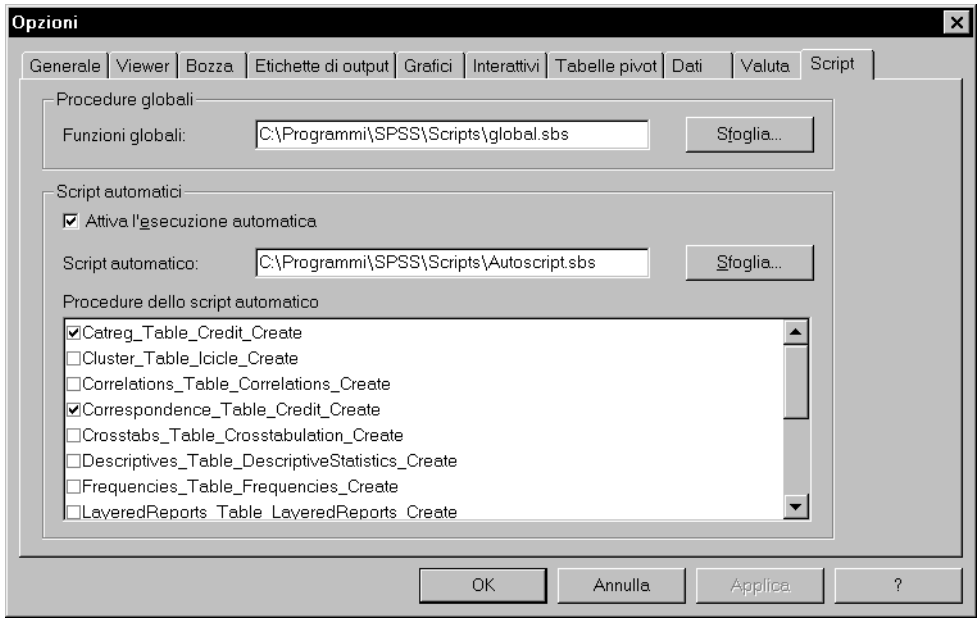
Nota: Questo elenco potrebbe non essere completo.

Script automatici

Gli script automatici vengono eseguiti automaticamente se avviati dalla creazione di un output specifico secondo una determinata procedura. Ad esempio, in SPSS è disponibile uno script che consente di rimuovere automaticamente la diagonale superiore e di evidenziare i coefficienti di correlazione inferiori a un certo livello di significatività ad ogni creazione di una tabella Correlazioni tramite la procedura Correlazioni bivariate.

Nella scheda Script della finestra di dialogo Opzioni (menu Modifica) sono visualizzati gli script automatici disponibili nel sistema. Ciascuno di tali script può essere attivato o disattivato.

Figura 47-2
Scheda Script della finestra di dialogo Opzioni



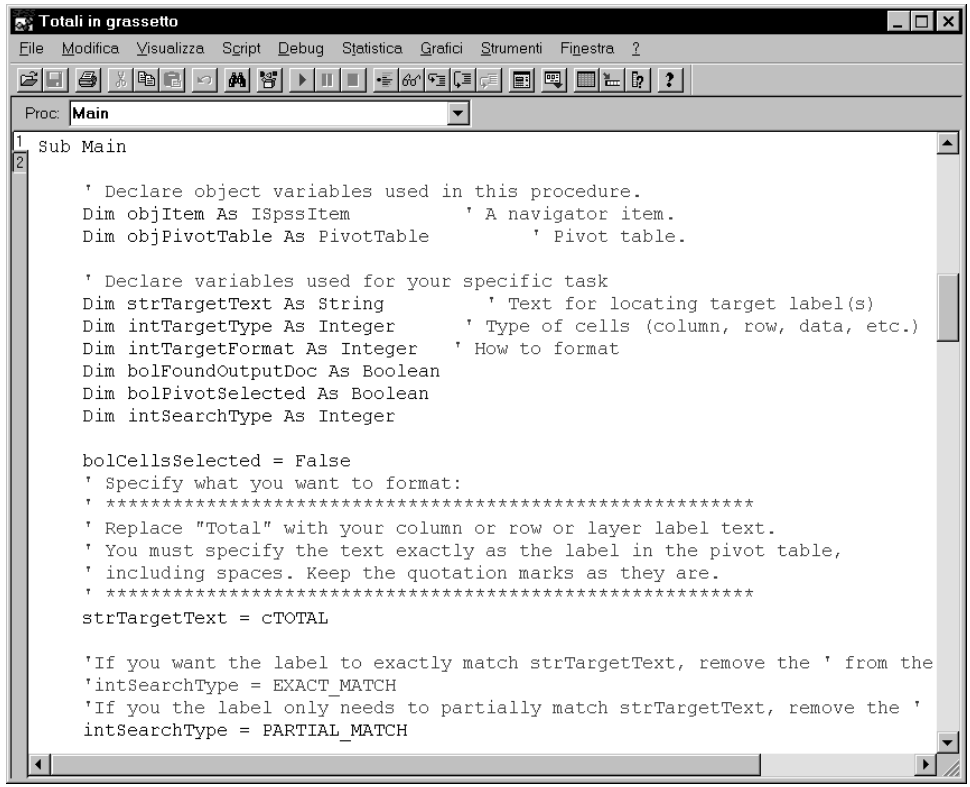
A ciascuna procedura e tipo di output corrisponde uno script automatico specifico. Uno script automatico per la formattazione di tabelle ANOVA create tramite una procedura ANOVA univariata non viene attivato da tabelle ANOVA create tramite altre procedure statistiche. Per queste tabelle ANOVA, che includono tutte codice simile, è tuttavia possibile usare procedure globali per la creazione di script automatici distinti. A ciascun tipo di output ottenuto con una stessa procedura è inoltre possibile associare uno script automatico distinto. Con la procedura Frequenze, ad esempio, si ottengono una tabella di frequenze e una tabella di statistiche, a ciascuna delle quali è possibile associare uno script automatico distinto.

Per ulteriori informazioni, vedere “Opzioni” in Capitolo 44 a pag. 629.

Creazione e modifica di script

Molti degli script disponibili in SPSS possono essere personalizzati in base alle specifiche esigenze. Esiste ad esempio uno script che consente di eliminare tutte le tabelle Note da un output designato. È possibile modificare con facilità questo script in modo da eliminare qualsiasi elemento e l'etichetta desiderata.

Figura 47-3
Modifica di uno script nella finestra dello script


Totali in grassetto
File Modifica Visualizza Script Debug Statistica Grafici Strumenti Finestra ?
Proc: Main
Sub Main
' Declare object variables used in this procedure.
Dim objItem As ISpssItem ' A navigator item.
Dim objPivotTable As PivotTable ' Pivot table.
' Declare variables used for your specific task
Dim strTargetText As String ' Text for locating target label(s)
Dim intTargetType As Integer ' Type of cells (column, row, data, etc.)
Dim intTargetFormat As Integer ' How to format
Dim bolFoundOutputDoc As Boolean
Dim bolPivotSelected As Boolean
Dim intSearchType As Integer
bolCellsSelected = False
' Specify what you want to format:
' **
' Replace "Total" with your column or row or layer label text.
' You must specify the text exactly as the label in the pivot table,
' including spaces. Keep the quotation marks as they are.
' **
strTargetText = cTOTAL
'If you want the label to exactly match strTargetText, remove the ' from the
'intSearchType = EXACT_MATCH
'If you the label only needs to partially match strTargetText, remove the '
intSearchType = PARTIAL_MATCH

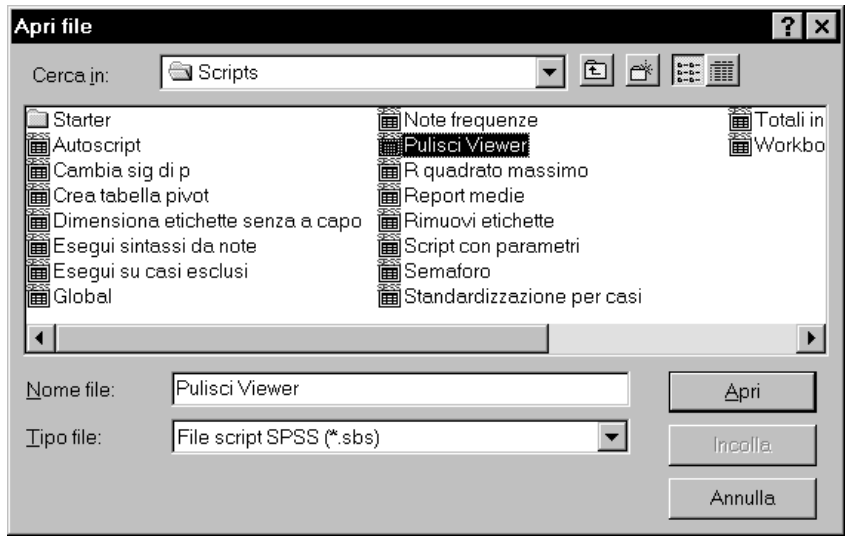
Se si desidera creare nuovi script, è possibile usare come punto di partenza uno degli script di avvio.

Per modificare uno script

- Dai menu, scegliere:

  File
  Apri
  Script...

Figura 47-4
Apertura di uno script



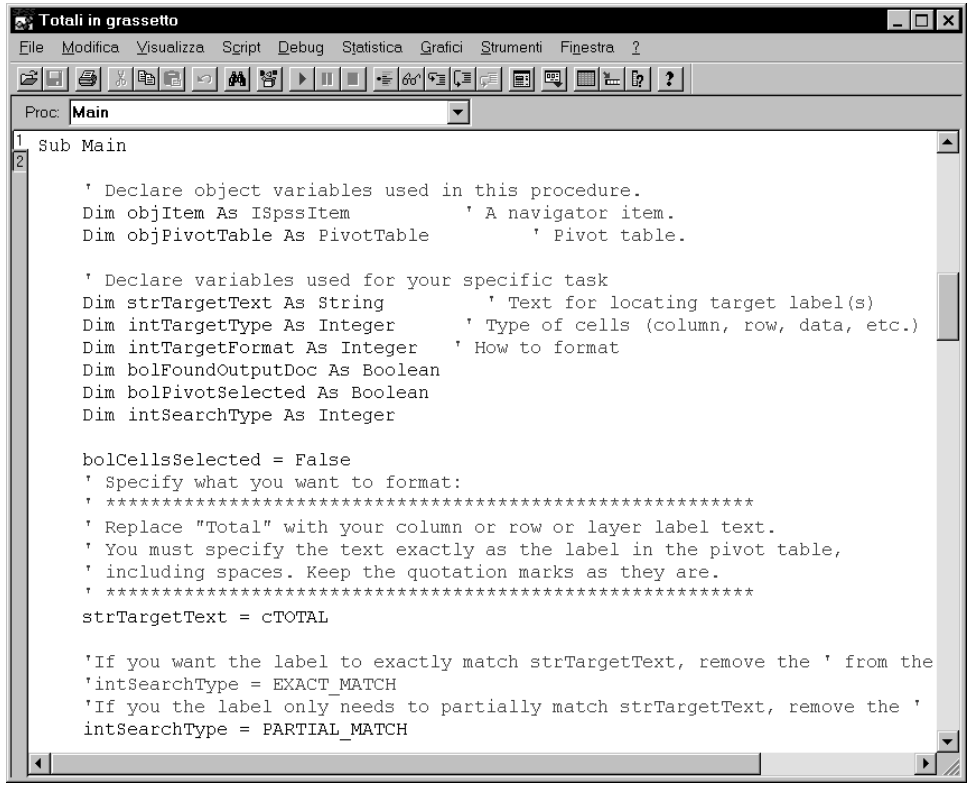
- Selezionare la cartella *Script*.
- Nella casella Tipo file selezionare File script SPSS (*.sbs).
- Selezionare lo script desiderato.

  Se si aprono più script, ciascuno di essi verrà aperto in una finestra distinta.

finestra dello script

La finestra dello script è un ambiente di programmazione con funzionalità complete che usa il linguaggio BASIC Sax. Include un editor di finestre di dialogo, un visualizzatore oggetti, funzioni di debug e un Aiuto contestuale.

Figura 47-5
finestra dello script


Totali in grassetto
File Modifica Visualizza Script Debug Statistica Grafici Strumenti Finestra ?
Proc: Main
Sub Main
' Declare object variables used in this procedure.
Dim objItem As ISpssItem ' A navigator item.
Dim objPivotTable As PivotTable ' Pivot table.
' Declare variables used for your specific task
Dim strTargetText As String ' Text for locating target label(s)
Dim intTargetType As Integer ' Type of cells (column, row, data, etc.)
Dim intTargetFormat As Integer ' How to format
Dim bolFoundOutputDoc As Boolean
Dim bolPivotSelected As Boolean
Dim intSearchType As Integer
bolCellsSelected = False
' Specify what you want to format:
' **
' Replace "Total" with your column or row or layer label text.
' You must specify the text exactly as the label in the pivot table,
' including spaces. Keep the quotation marks as they are.
' **
strTargetText = cTOTAL
'If you want the label to exactly match strTargetText, remove the ' from the
'intSearchType = EXACT_MATCH
'If you the label only needs to partially match strTargetText, remove the '
intSearchType = PARTIAL_MATCH

- Spostando il cursore, viene visualizzato il nome della procedura corrente nella parte superiore della finestra.
- I termini in blu corrispondono a parole riservate del linguaggio BASIC, ad esempio `Sub`, `End Sub` e `Dim`). Per accedere all'Aiuto contestuale relativo a questi termini, è possibile fare clic sui termini e premere F1.
- I termini in magenta corrispondono a oggetti, proprietà e metodi. Anche in questo caso è possibile visualizzare informazioni dell'Aiuto facendo clic sui termini e premendo F1. Ciò è possibile tuttavia solo se i termini sono visualizzati in colore magenta e si trovano in istruzioni valide. Facendo clic sul nome di un oggetto in un commento, non verranno visualizzate informazioni sugli oggetti SPSS, ma l'Aiuto relativo al linguaggio BASIC Sax.

- I commenti sono in verde.
- Premere F2 in qualsiasi momento per visualizzare il Visualizzatore oggetti, in cui sono elencati gli oggetti, le proprietà e i metodi.

Proprietà dell'Editor degli script (finestra dello script)

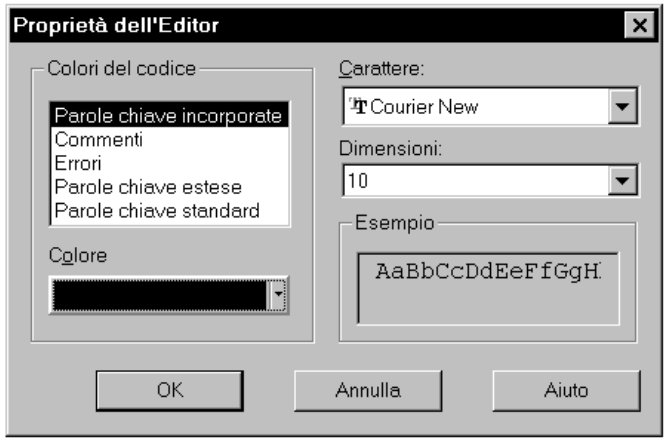
Nella finestra dello script gli elementi del codice sono contraddistinti da diversi colori e possono pertanto essere individuati in modo rapido. Per impostazione predefinita, i commenti sono in verde, i termini BASIC Sax sono in blu e i nomi di oggetti, proprietà e metodi validi sono in magenta. È tuttavia possibile impostare colori diversi per tali elementi e modificare le dimensioni e il carattere di tutto il testo.

Per impostare le proprietà dell'Editor degli script

▶ Dai menu, scegliere:

Script
  Proprietà editor…

Figura 47-6
Finestra di dialogo Proprietà dell'editor

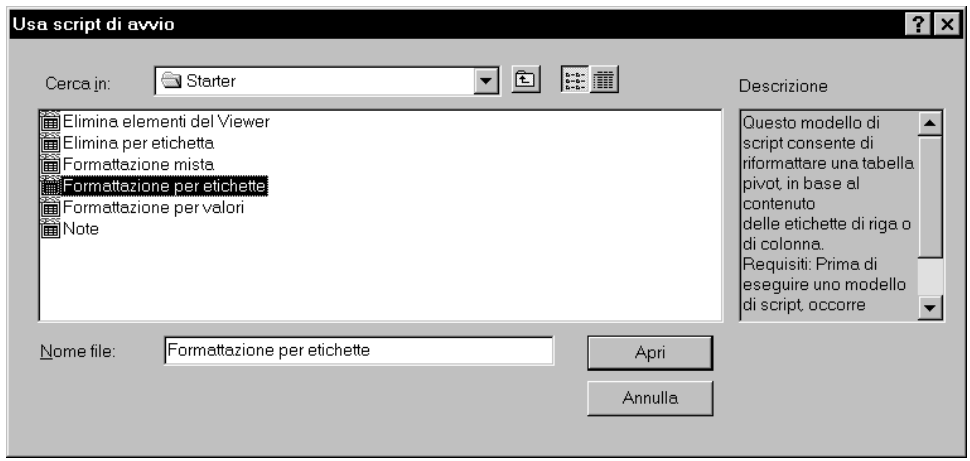


▶ Per modificare il colore di un tipo di elemento del codice, selezionare l'elemento e quindi il colore desiderato dalla tavolozza.

Script di avvio

Quando si crea un nuovo script, è possibile scegliere come punto di partenza uno degli script di avvio disponibili.

Figura 47-7
Finestra di dialogo Usa script di avvio


Usa script di avvio
Cerca in:
Starter
Descrizione
Elimina elementi del Viewer
Elimina per etichetta
Formattazione mista
Formattazione per etichette
Formattazione per valori
Note
Questo modello di script consente di riformattare una tabella pivot, in base al contenuto delle etichette di riga o di colonna. Requisiti: Prima di eseguire uno modello di script, occorre
Nome file:
Formattazione per etichette
Apri
Annulla

Ciascuno script di avvio include codice per una o più procedure usate di frequente con suggerimenti sulla personalizzazione dello script in base alle specifiche esigenze.

Elimina per etichetta. Consente di eliminare righe o colonne di una tabella pivot in base al contenuto di RowLabels o ColumnLabels. Questo script funziona correttamente se è selezionata l'opzione Nascondi righe e colonne vuote della finestra di dialogo Proprietà della tabella.

Elimina elementi del Viewer. Consente di eliminare gli elementi del Viewer in base a determinati criteri.

Piè di pagina. Consente di riformattare le didascalie di una tabella pivot, di modificarne il testo e di aggiungere nuove didascalie.

Formattazione per etichette. Consente di riformattare una tabella pivot in base alle etichette di riga, di colonna o di strato.

Formattazione per valori. Consente di riformattare una tabella pivot in base al valore delle celle di dati o a una combinazione di celle ed etichette.

Formattazione mista. Consente di riformattare o modificare il testo del titolo, l'intestazione e il testo che compare in alto a sinistra in una tabella pivot.

È inoltre possibile usare come script di avvio uno degli altri script di SPSS, che tuttavia non sono altrettanto semplici da personalizzare. È sufficiente aprire lo script desiderato e salvarlo con un nome di file diverso.

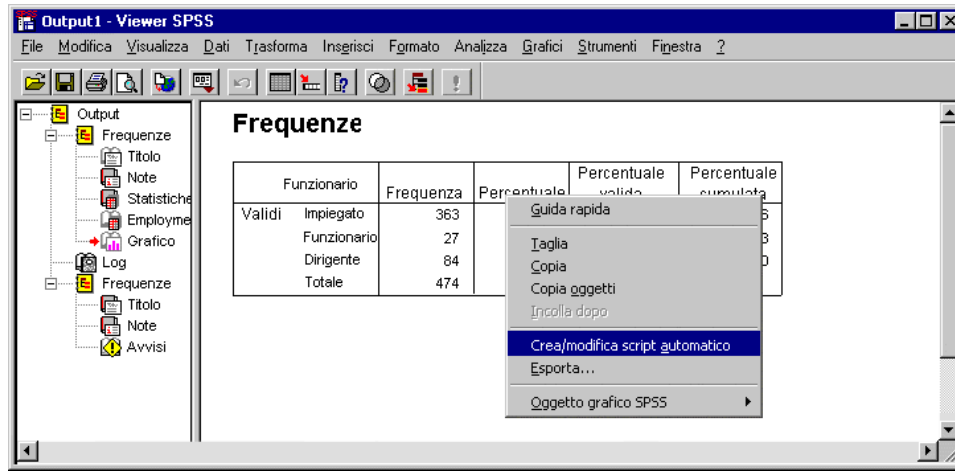
Per creare uno script

- Dai menu, scegliere:
  Nuovo
  Script…
- Selezionare uno script di avvio, se si desidera usarne uno.
- In caso contrario, fare clic su Annulla.

Creazione di script automatici

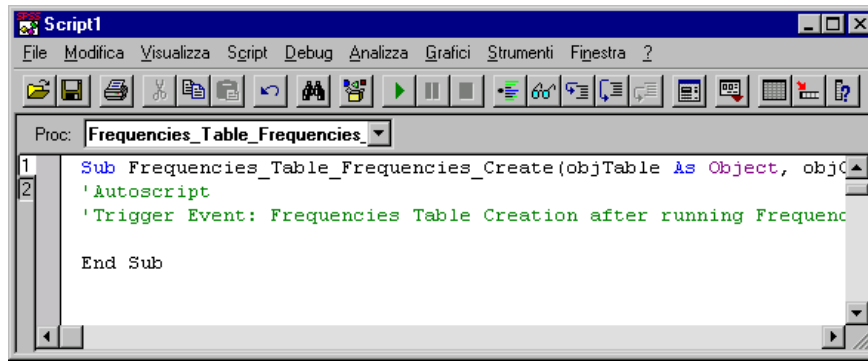
Per creare uno script automatico, è necessario partire dall'oggetto di output che si desidera usare come elemento di attivazione. Ad esempio per creare uno script automatico che venga eseguito ogni volta che viene creata una tabella di frequenza, creare tale tabella seguendo la normale procedura e fare clic sulla tabella nel Viewer per selezionarla. Sarà quindi possibile fare clic con il pulsante destro oppure usare il menu Strumenti per creare un nuovo script automatico che verrà eseguito ogni volta che viene creato quel tipo di tabella.

Figura 47-8
Creazione di nuovi script automatici



Per impostazione predefinita. i nuovi script automatici creati vengono aggiunti come nuova procedura nello script automatico corrente, *autscript.sbs*. Il nome della procedura fa riferimento all'evento che agisce da elemento di attivazione. Se si crea ad esempio uno script automatico che si avvia ogni volta che tramite la procedura Esplora viene creata una tabella Descrittive, il nome della procedura sarà `Explore_Table_Descriptives_Create`.

Figura 47-9
Nuova procedura di script automatico visualizzata nella finestra dello script


Script1
File Modifica Visualizza Script Debug Analizza Grafici Strumenti Finestra ?
Proc: Frequencies_Table_Frequencies
Sub Frequencies_Table_Frequencies_Create(objTable As Object, obj
'Autoscript
'Trigger Event: Frequencies Table Creation after running Frequenc
End Sub

Ciò semplifica la creazione di script automatici in quanto non è necessario scrivere il codice per ottenere l'oggetto su cui eseguire operazioni. Gli script automatici devono tuttavia essere specifici per un output e una procedura statistica specifici.

Per creare uno script automatico

- Nel Viewer selezionare l'oggetto che si desidera funzioni come elemento di attivazione.
- Dai menu, scegliere:

  Strumenti
  Crea/modifica script automatico…

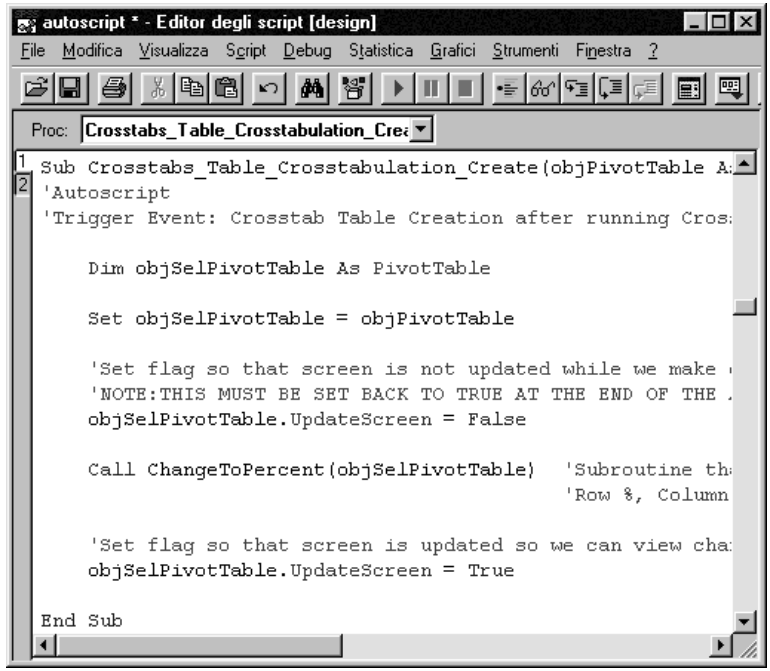
  Se per l'oggetto selezionato non esiste alcuno script automatico, ne verrà automaticamente creato uno. Se lo script è disponibile, verrà visualizzato.
- Inserire il codice.
- Scegliere Opzioni dal menu Modifica per attivare o disattivare lo script automatico.

Eventi che attivano gli script automatici

Il nome della procedura dello script automatico fa riferimento all'evento che agisce da elemento di attivazione. Gli script automatici possono essere attivati dai seguenti eventi:

Creazione di una tabella pivot. Il nome della procedura fa riferimento sia al tipo di tabella sia alla procedura con cui la tabella è stata creata, ad esempio `Correlations_Table_Correlations_Create`.

Figura 47-10
Procedura di script automatico per la tabella Correlazioni


autoscript * - Editor degli script [design]
File Modifica Visualizza Script Debug Statistica Grafici Strumenti Finestra ?
Proc: Crosstabs_Table_Crosstabulation_Crea
Sub Crosstabs_Table_Crosstabulation_Create(objPivotTable A:
'Autoscript
'Trigger Event: Crosstab Table Creation after running Cros:
Dim objSelPivotTable As PivotTable
Set objSelPivotTable = objPivotTable
'Set flag so that screen is not updated while we make
'NOTE:THIS MUST BE SET BACK TO TRUE AT THE END OF THE
objSelPivotTable.UpdateScreen = False
Call ChangeToPercent(objSelPivotTable) 'Subroutine th
'Row %, Column
'Set flag so that screen is updated so we can view cha
objSelPivotTable.UpdateScreen = True
End Sub

Creazione del titolo. Riferimento alla procedura statistica con cui è stata creata: `Correlations_Title_Create`.

Creazione di note. Riferimento alla procedura con cui sono state create: `Correlations_Notes_Create`.

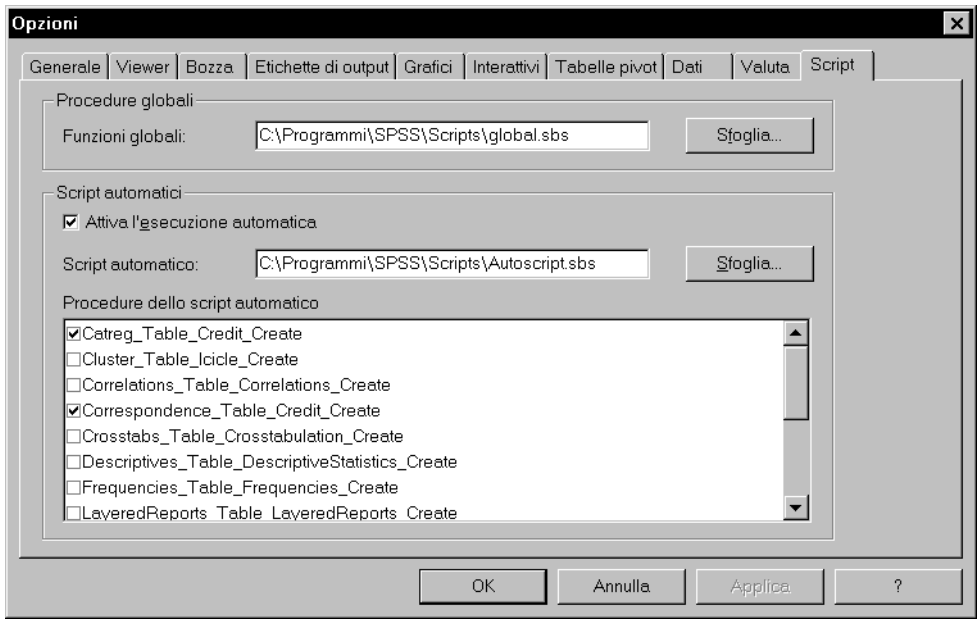
Creazione di avvisi. Riferimento alla procedura con cui sono stati creati.

Si possono inoltre usare gli script per attivare script automatici in modo indiretto. È possibile, ad esempio, scrivere uno script che richiami la procedura Correlazioni, la quale a sua volta attiva lo script automatico registrato nella tabella Correlazioni risultante.

Script automatico

Tutti gli script automatici, a differenza degli altri script che vengono salvati separatamente in file distinti, vengono salvati in un unico file. In questo file vengono inoltre aggiunti tutti i nuovi script automatici creati. Il nome del file degli script automatici corrente è visualizzato nella scheda Script della finestra di dialogo Opzioni (menu Modifica).

Figura 47-11
Procedure dello script automatico visualizzate nella finestra di dialogo Opzioni



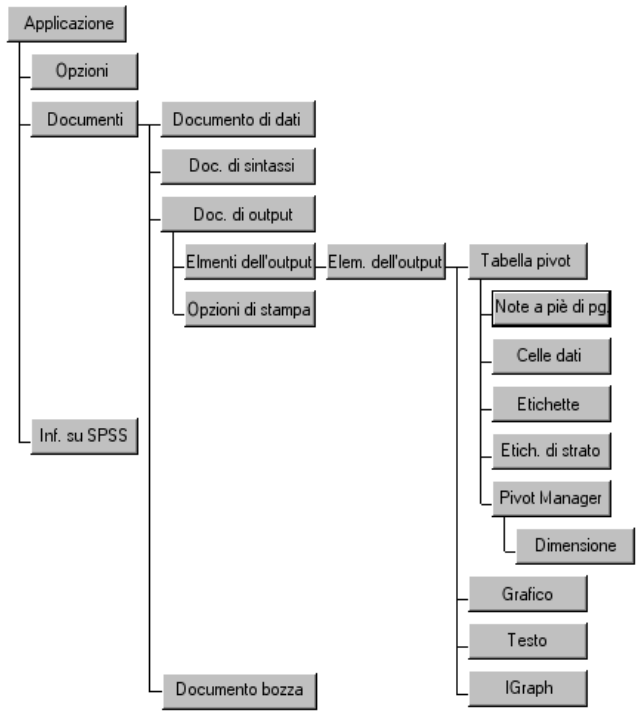
Nella finestra di dialogo Opzioni sono inoltre visualizzati tutti gli script automatici inclusi nel file selezionato. Ciascuno di tali script può essere attivato o disattivato.

Il file degli script automatici predefinito è *autscript.sbs*. È possibile specificare un file degli script automatici diverso ma non può esserne attivo più di uno contemporaneamente.

Modalità di funzionamento degli script

L'esecuzione di script consiste nell'esecuzione di operazioni sugli oggetti tramite proprietà e metodi. Le tabelle pivot, ad esempio, sono una classe di oggetti. Con gli oggetti di questa classe, è possibile usare il metodo SelectTable per selezionare tutti gli elementi della tabella e la proprietà TextColor per modificare il colore del testo selezionato. A ciascuna classe dell'oggetto è associato un insieme specifico di proprietà e metodi. L'insieme di tutte le classi, o tipi, di oggetti SPSS è definito libreria dei tipi SPSS.

Figura 47-12
Visualizzazione della struttura gerarchica degli oggetti



L'uso di oggetti è un processo a due fasi. Occorre prima creare un riferimento all'oggetto, ovvero *ottenere* l'oggetto. Quindi, è possibile usare proprietà e metodi per l'esecuzione di operazioni specifiche. Per ottenere gli oggetti è necessario spostarsi nella gerarchia degli oggetti SPSS usando ad ogni livello le proprietà o i metodi degli

oggetti di livello superiore in modo da ottenere gli oggetti di tutti i livelli inferiori. Per ottenere un oggetto tabella pivot, ad esempio, è necessario ottenere innanzitutto l'output contenente la tabella e quindi gli elementi dell'output.

Ciascun oggetto viene memorizzato in una variabile. È importante ricordare che nella variabile viene semplicemente memorizzato un riferimento all'oggetto. Una delle prime operazioni da eseguire durante la creazione di uno script consiste nel dichiarare le variabili relative agli oggetti necessari.

Suggerimento: Se non è chiara la modalità di funzionamento di SPSS, risulterà difficile capire la modalità di funzionamento degli script. Prima di iniziare a scrivere uno script, provare ad eseguire più volte l'operazione richiesta usando il mouse. Ad ogni passaggio analizzare gli oggetti su cui si sta intervenendo e le proprietà degli oggetti che si stanno modificando.

Dichiarazione di variabili (script)

Sebbene non sia sempre richiesto, è consigliabile dichiarare tutte le variabili prima di usarle. Questa operazione viene in genere eseguita con istruzioni di dichiarazione `Dim`:

```
Dim objOutputDoc As ISpssOutputDoc
Dim objPivotTable As PivotTable
Dim intType As Integer
Dim strLabel As String
```

In ciascuna dichiarazione vengono specificati il nome e il tipo della variabile. Con la prima delle dichiarazioni riportate sopra, ad esempio, viene creata la variabile oggetto `objOutputDoc` che viene assegnata alla classe di oggetti `ISpssOutputDoc`. Alla variabile non è tuttavia associato alcun valore in quanto non è ancora stata impostata su un documento di output specifico. Nell'istruzione viene semplicemente dichiarato che la variabile esiste. Questo processo è stato a volte descritto come "ridenominazione degli oggetti che si desidera usare".

Convenzioni di denominazione delle variabili. Per convenzione, il nome di ciascuna variabile indica il tipo di variabile. I nomi delle variabili oggetto iniziano con `obj`, le variabili di intero con `int` e le variabili stringa con `str`. Sebbene sia possibile

assegnare alle variabili qualsiasi nome, seguendo queste convenzioni risulterà più semplice capire il codice.

Classi di oggetti di SPSS. `ISpssOutputDoc` e `PivotTable` sono nomi di classi di oggetti di SPSS. Ciascuna classe rappresenta un tipo di oggetto creato in SPSS, ad esempio un output o una tabella pivot. A ciascuna classe dell'oggetto è associato un insieme specifico di proprietà e metodi. L'insieme di tutte le classi, o tipi, di oggetti SPSS viene definito libreria dei tipi SPSS.

Convenzioni di denominazione per le classi e le variabili oggetto SPSS

Di seguito sono elencati i nomi di variabile usati negli script di esempio inclusi in SPSS. È consigliabile usare questi nomi in tutti gli script. Ad eccezione delle tabelle pivot, i nomi delle classi di oggetti iniziano con `ISpss`.

Oggetto	Tipo o classe	Nome di variabile
Applicazione SPSS	`IspssApp`	`objSpssApp`—La variabile è globale e di conseguenza non è necessario dichiararla
Opzioni SPSS	`IspssOptions`	`objSpssOptions`
File di informazioni SPSS	`IspssInfo`	`objSpssInfo`
Documenti	`ISpssDocuments`	`objDocuments`
Documento di dati	`ISpssDataDoc`	`objDataDoc`
Documento di sintassi	`ISpssSyntaxDoc`	`objSyntaxDoc`
Documento Viewer	`ISpssOutputDoc`	`objOutputDoc`
Opzioni di stampa	`ISpssPrintOptions`	`objPrintOptions`
Insieme degli elementi di output	`ISpssItems`	`objOutputItems`
Elemento di output	`ISpssItem`	`objOutputItem`
Grafici	`ISpssChart`	`objSPSSChart`
Testo	`ISpssRtf`	`objSPSSText`
Tabella pivot	`PivotTable`	`objPivotTable`
Note a piè di pagina	`ISpssFootnotes`	`objFootnotes`
Celle di dati	`ISpssDataCells`	`objDataCells`
Etichette di strato	`ISpssLayerLabels`	`objLayerLabels`

Oggetto	Tipo o classe	Nome di variabile
Etichette di colonna	ISpssLabels	objColumnLabels
Etichette di riga	ISpssLabels	objRowLabels
Pivot manager	ISpssPivotMgr	objPivotMgr
Dimensione	ISpssDimension	objDimension

Per ottenere oggetti di automazione SPSS (script)

Ottenere un oggetto significa creare un riferimento all'oggetto in modo da poter usare proprietà e metodi per l'esecuzione di operazioni specifiche. Ciascun riferimento all'oggetto viene memorizzato in una variabile. Per ottenere un oggetto, è innanzitutto necessario dichiarare una variabile oggetto della classe appropriata e impostare quindi la variabile sull'oggetto specifico. Per ottenere, ad esempio, l'output designato:

```
Dim objOutputDoc As ISpssOutputDoc
Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc
```

si utilizzano le proprietà e i metodi degli oggetti del livello superiore nella gerarchia degli oggetti SPSS per ottenere gli oggetti di livello inferiore. La seconda istruzione consente di ottenere l'output designato usando il metodo `GetDesignatedOutputDoc` associato all'oggetto Applicazione SPSS, ovvero all'oggetto di primo livello. In modo analogo, per ottenere un oggetto tabella pivot, è innanzitutto necessario ottenere l'output contenente la tabella pivot, quindi l'insieme degli elementi dell'output e così via.

Esempio: come ottenere un elemento di output

Lo script riportato di seguito consente di ottenere e attivare il terzo elemento di output incluso nell'output designato. Se tale elemento non è un oggetto OLE, verrà generato un errore.

```
Sub Main

Dim objOutputDoc As ISpssOutputDoc'declare object variables
Dim objOutputItems As ISpssItems
Dim objOutputItem As ISpssItem
```

```
Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc'get reference to designated output doc
Set objOutputItems = objOutputDoc.Items() 'get collection of items in doc
Set objOutputItem = objOutputItems.GetItem(2) 'get third output item
'(item numbers start at 0 so "2" gets third)

objOutputItem.Activate 'activate output item

End sub
```

Esempio: Per ottenere la prima tabella pivot

Lo script riportato di seguito consente di ottenere e attivare la prima tabella pivot inclusa nell'output designato.

```
Sub Main

Dim objOutputDoc As ISpssOutputDoc 'declare object variables
Dim objOutputItems As ISpssItems
Dim objOutputItem As ISpssItem
Dim objPivotTable As PivotTable

Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc'get reference to designated output doc
Set objOutputItems = objOutputDoc.Items()'get collection of items in doc

Dim intItemCount As Integer'number of output items
Dim intItemType As Integer'type of item (defined by SpssType property)

intItemCount = objOutputItems.Count()'get number of output items
    For index = 0 To intItemCount'loop through output items
   Set objOutputItem = objOutputItems.GetItem(index)'get current item
    intItemType = objOutputItem.SPSSType()'get type of current item
    If intItemType = SPSSPivot Then
        Set objPivotTable = objOutputItem.Activate()'if item is a pivot table, activate it
        Exit For
   End If
Next index

End sub
```

Gli esempi sono inoltre disponibili nell'Aiuto in linea. Gli esempi possono essere eseguiti copiando il codice dall'Aiuto e incollandolo nella finestra dello script.

Proprietà e metodi (script)

Analogamente agli oggetti reali, anche agli oggetti di automazione OLE hanno funzioni e usi diversi. Nella terminologia della programmazione, le funzioni sono denominate proprietà e gli usi, metodi. A ciascuna classe dell'oggetto è associato un gruppo specifico di metodi e proprietà che determinano il tipo di operazioni eseguibili con l'oggetto.

Oggetto	Proprietà	Metodo
Matita (oggetto reale)	Durezza Colore	Scrittura Cancellazione
Tabella pivot (SPSS)	TextFont DataCellWidths CaptionText	SelectTable ClearSelection HideFootnotes

Esempio: uso delle proprietà (script)

Le proprietà consentono di impostare o restituire attributi degli oggetti, quali il colore o la larghezza delle celle. Quando una proprietà è visualizzata a sinistra di un segno di uguale, significa che si sta scrivendo in tale proprietà. La riga seguente, ad esempio, consente di impostare la didascalia di una tabella pivot attiva (`objPivotTable`) su `"Risultati di Anita"`:

```
objPivotTable.CaptionText = "Risultati di Anita"
```

Quando una proprietà è visualizzata a destra del segno di uguale, significa che si sta leggendo tale proprietà. La riga seguente, ad esempio, indica che si sta leggendo e salvando in una variabile la didascalia della tabella pivot attiva:

```
strFontName = objPivotTable.CaptionText
```

Esempio: uso dei metodi (script)

I metodi consentono di eseguire operazioni sugli oggetti, ad esempio selezionare tutti gli elementi inclusi in una tabella:

```
objPivotTable.SelectTable
```

o rimuovere una selezione:

```
objPivotTable.ClearSelection
```

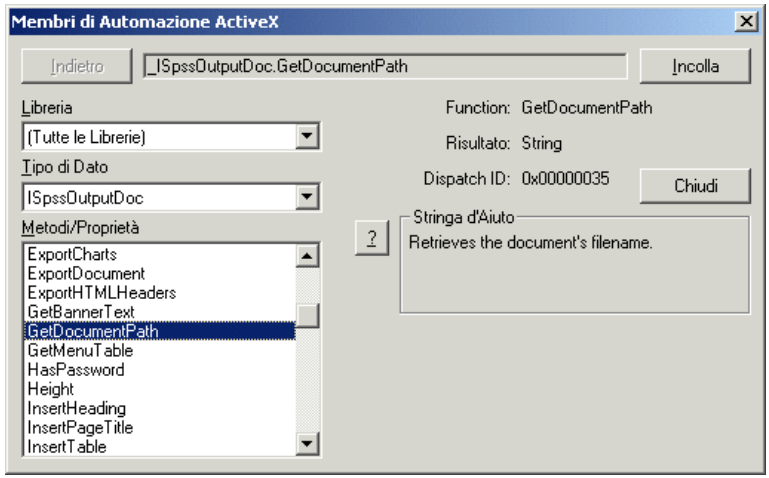
Alcuni metodi restituiscono un altro oggetto. Tali metodi risultano estremamente importanti per spostarsi nella gerarchia degli oggetti di SPSS. Il metodo `GetDesignatedOutputDoc`, ad esempio, restituisce l'output designato consentendo di accedere ai relativi elementi:

```
Set objOutputDoc = objSpssApp.GetDesignatedOutputDoc
Set objItems = objOutputDoc.Items
```

Visualizzatore oggetti

Nel Visualizzatore oggetti sono elencate tutte le classi di oggetti SPSS e i corrispondenti metodi e proprietà. È inoltre possibile accedere alle informazioni dell'Aiuto su proprietà e metodi singoli e incollare nello script le proprietà e i metodi selezionati.

Figura 47-13
Visualizzatore oggetti



Per usare il Visualizzatore oggetti

- Nella finestra dello script scegliere:

  Debug
  Visualizzatore oggetti...

- Selezionare una classe di oggetti dall'elenco Tipo di dati per visualizzare metodi e proprietà corrispondenti.

- Selezionare le proprietà e i metodi di cui si desidera visualizzare l'Aiuto contestuale o da incollare nello script.

Nuova procedura (script)

Una procedura è una sequenza di istruzioni eseguite come singola unità. L'organizzazione del codice in procedure rende più semplice la gestione e il riutilizzo del codice stesso. Gli script devono includere almeno una procedura, la procedura `Main`, ma spesso ne includono diverse. Oltre alle chiamate di procedure che eseguono la maggior parte delle operazioni, la procedura `Main` può includere alcune istruzioni.

Figura 47-14
Finestra di dialogo Nuova procedura

Esistono procedure e funzioni. Una procedura inizia con un'istruzione che specifica il tipo e il nome della procedura, ad esempio `Sub Main` o `Function DialogMonitor( )` e termina con l'istruzione End appropriata, ovvero `End Sub` o `End Function`.

Mentre si scorre la finestra dello script, il nome della procedura corrente viene visualizzato nella parte superiore. All'interno di uno script, è possibile chiamare qualsiasi procedura tutte le volte che si desidera. È inoltre possibile chiamare qualsiasi procedura nel file di script globale, che consente di condividere le procedure tra script.

Per aggiungere una nuova procedura in uno script

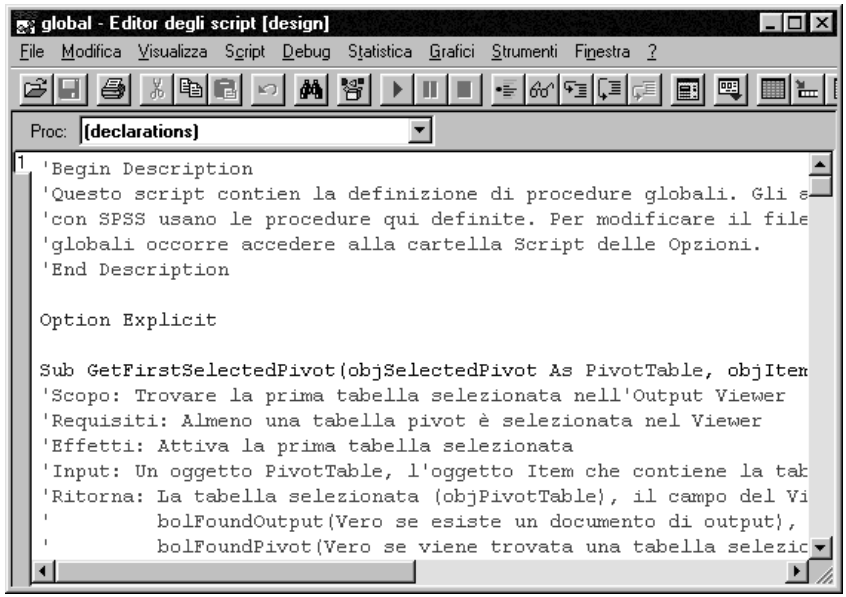
- Dai menu, scegliere:
  Script
  Nuova procedura...
- Inserire un nome per la procedura.
- Selezionare Procedura o Funzione.

In alternativa è possibile creare una nuova procedura inserendo le istruzioni che la definiscono direttamente nello script.

Procedure globali (script)

Se si desidera usare una procedura o una funzione in più script diversi, è possibile aggiungerla nel file di script globale. Le procedure incluse nel file di script globale possono essere chiamate da tutti gli altri script.

Figura 47-15
File di script globale


global - Editor degli script [design]
File Modifica Visualizza Script Debug Statistica Grafici Strumenti Finestra ?
Proc: (declarations)
'Begin Description
'Questo script contien la definizione di procedure globali. Gli s
'con SPSS usano le procedure qui definite. Per modificare il file
'globali occorre accedere alla cartella Script delle Opzioni.
'End Description
Option Explicit
Sub GetFirstSelectedPivot(objSelectedPivot As PivotTable, objItem
'Scopo: Trovare la prima tabella selezionata nell'Output Viewer
'Requisiti: Almeno una tabella pivot è selezionata nel Viewer
'Effetti: Attiva la prima tabella selezionata
'Input: Un oggetto PivotTable, l'oggetto Item che contiene la tab
'Ritorna: La tabella selezionata (objPivotTable), il campo del Vi
' bolFoundOutput(Vero se esiste un documento di output),
' bolFoundPivot(Vero se viene trovata una tabella selezio

Il file di script globale predefinito è *global.sbs*, a cui è possibile aggiungere ulteriori procedure. È inoltre possibile impostare un file di script globale diverso nella scheda Script della finestra di dialogo Opzioni (menu Modifica), ma un solo file alla volta può essere attivo come file globale. Ciò significa che se si crea e si imposta come file globale un nuovo file, le procedure e le funzioni incluse in *global.sbs* non saranno più disponibili.

È possibile visualizzare il file di script globale in qualsiasi finestra dello script facendo clic sulla scheda n. 2 a sinistra della finestra immediatamente sotto la barra degli strumenti, ma è possibile modificarlo in una sola finestra alla volta.

Le procedure globali devono essere richiamate da altre procedure di script. Non è possibile eseguire uno script globale direttamente dal menu Strumenti o da una finestra dello script.

Aggiunta di descrizioni agli script

È possibile aggiungere una descrizione da visualizzare nelle finestre di dialogo Esegui script e Usa script di avvio. È sufficiente aggiungere un commento alla prima riga dello script che inizia con l'istruzione `Begin Description` seguita dal commento desiderato di una o più righe e dall'istruzione `End Description`. Ad esempio:

```
'Begin Description
'Questo script cambia "Sig." in "p=" nelle etichette delle colonne delle tabelle pivot.
'Requisito: è necessario selezionare la tabella pivot da modificare.
'End Description
```

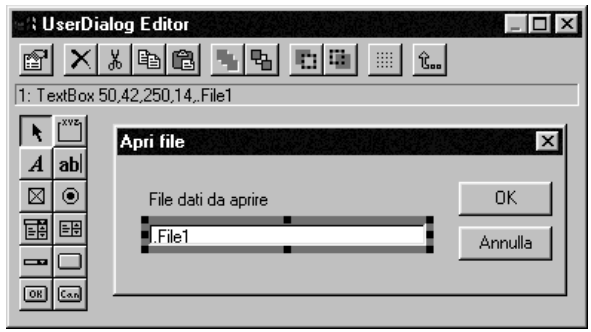
La descrizione deve essere formattata come commento, ovvero ciascuna riga deve iniziare con un apostrofo.

Script per la creazione di finestre di dialogo personalizzate

Per implementare una finestra di dialogo è necessario seguire la seguente procedura: innanzitutto creare una finestra di dialogo tramite UserDialog Editor, quindi creare una funzione di controllo (`DialogFunc`) per il monitoraggio della finestra di dialogo e la definizione del relativo funzionamento.

La stessa finestra di dialogo è definita da un blocco `Begin Dialog...End Dialog`. Non è necessario digitare questo codice direttamente.—UserDialog Editor consente di definire graficamente e in modo semplice le finestre di dialogo.

Figura 47-16
Creazione di finestre di dialogo in UserDialog Editor



Nell'editor è inizialmente visualizzato il modulo di una finestra di dialogo vuoto. È possibile aggiungere controlli, quali pulsanti di opzione e caselle di controllo selezionando lo strumento appropriato e trascinando con il mouse. Per visualizzare una descrizione di uno strumento, posizionare il cursore sullo strumento desiderato. È inoltre possibile trascinare i lati e gli angoli di una finestra di dialogo per ridimensionarla. Dopo aver aggiunto un controllo, fare clic con il pulsante destro del mouse sul controllo per impostarne le proprietà.

Funzione di controllo della finestra di dialogo. Per creare una funzione di controllo, fare clic con il pulsante destro sul modulo della finestra di dialogo, verificando che non sia selezionato alcun controllo, quindi inserire un nome per la funzione nel campo DialogFunc. Le istruzioni che definiscono la funzione verranno aggiunte allo script. Per definirne il comportamento, tuttavia, ciascuna funzione dovrà essere modificata in modo manuale.

Dopo aver completato l'operazione, fare clic sul pulsante Salva ed esci, all'estremità destra della barra degli strumenti, per aggiungere il codice relativo alla finestra di dialogo nello script.

Per creare una finestra di dialogo personalizzata

- Nella finestra dello script, fare clic nel punto dello script in cui si desidera inserire il codice della finestra di dialogo.
- Dai menu, scegliere:

  Script
  Editor finestre di dialogo...

- Selezionare gli strumenti desiderati e trascinarli nel nuovo modulo della finestra di dialogo per aggiungere i controlli, quali pulsanti e caselle di controllo.
- Ridimensionare la finestra di dialogo trascinando i quadratini di ridimensionamento visualizzati sui lati e sugli angoli.
- Fare clic sul modulo con il pulsante destro del mouse, controllando che non sia selezionato alcun controllo, e inserire un nome per la funzione di controllo nel campo DialogFunc.
- Dopo aver completato l'operazione, fare clic sul pulsante Salva ed esci, all'estremità destra della barra degli strumenti.

Per definire il comportamento della finestra di dialogo, è necessario modificare la funzione di controllo in modo manuale.

Funzioni di controllo delle finestre di dialogo (script)

Una funzione di controllo definisce il comportamento di una finestra di dialogo per un numero di casi specifico. La forma generica della funzione è la seguente:

```
Function DialogFunc(strDlgItem as String, intAction as Integer, intSuppValue as Integer)
  Select Case intAction
    Case 1 ' dialog box initialization
     ...    'statements to execute when dialog box is initialized
    Case 2 ' value changing or button pressed
     ...    'statements...
    Case 3 ' TextBox or ComboBox text changed ...
    Case 4 ' focus changed ...
    Case 5 ' idle ...
  End Select
End Function
```

Parametri. La funzione deve essere in grado di passare tre parametri: una stringa `strDlgItem` e due valori interi `intAction` e `intSuppValue`. I parametri sono valori scambiati tra la funzione e la finestra di dialogo, diversi a seconda dell'azione da eseguire.

Quando, ad esempio, un utente fa clic su un controllo della finestra di dialogo, il nome del controllo viene passato alla funzione come `strDlgItem` (il nome del campo è specificato nella definizione della finestra di dialogo). Il secondo parametro, `intAction`, è un valore numerico che indica l'azione che è stata eseguita nella finestra di dialogo. Il terzo parametro consente in determinati casi di aggiungere ulteriori informazioni. Nella definizione della funzione è necessario includere i tre parametri, anche se non sono tutti usati.

Select Case intAction. Il valore di `intAction` indica l'azione che è stata eseguita nella finestra di dialogo. Ad esempio, l'inizializzazione della finestra di dialogo, `intAction` = 1. Se l'utente preme un pulsante, `intAction` viene modificato in 2 e così via. Sono disponibili cinque diverse azioni ed è possibile specificare le istruzioni eseguite per ciascuna azione, come indicato di seguito. Non è necessario specificare

tutti e cinque i possibili casi, —ma solo quelli appropriati. Se, ad esempio, si desidera che all'inizializzazione non venga eseguita alcuna istruzione, omettere Case 1.

- **Case intAction = 1.** Consente di specificare istruzioni da eseguire durante l'inizializzazione della finestra di dialogo. È possibile, ad esempio, disattivare uno o più controlli o aggiungere un segnale acustico. La stringa `strDlgItem` è una stringa nulla, mentre`intSuppValue` è pari a 0.
- **Case 2.** Viene eseguito quando viene premuto un pulsante oppure quando un valore cambia in uno dei seguenti controlli: `CheckBox`, `DropListBox`, `ListBox` o `OptionGroup`. Se viene premuto un pulsante, `strDlgItem` è il pulsante, `intSuppValue` è privo di significato ed è necessario impostare `DialogFunc` = True per evitare la chiusura della finestra. Nel caso della modifica di un valore, `strDlgItem` corrisponde all'elemento di cui è stato modificato il valore e `intSuppValue` rappresenta il nuovo valore.
- **Case 3.** Viene eseguito quando un valore viene modificato in un controllo `TextBox` o `ComboBox`. La stringa `strDlgItem` corrisponde al controllo di cui è stato modificato il testo e da cui viene spostato lo stato attivo, mentre `intSuppValue` corrisponde al numero di caratteri.
- **Case 4.** Viene eseguito quando lo stato attivo viene spostato su un altro elemento della finestra di dialogo. La stringa `strDlgItem` rappresenta l'elemento su cui viene spostato lo stato attivo, mentre `intSuppValue` rappresenta l'elemento da cui lo stato attivo viene spostato (il primo elemento è 0, il secondo è 1 e così via).
- **Case 5.** Non viene eseguita alcuna operazione. La stringa `strDlgItem` rappresenta una stringa nulla, mentre `intSuppValue` è pari a 0. Per continuare a ricevere questo tipo di azioni, impostare `DialogFunc` = True.

Per ulteriori informazioni, vedere gli esempi e il prototipo di `DialogFunc` nella guida del linguaggio BASIC Sax.

Esempio: Script per la creazione di una semplice finestra di dialogo

Lo script riportato di seguito consente di creare una semplice finestra di dialogo per l'apertura di un file di dati. Vedere le sezioni correlate per ulteriori informazioni sulla procedura `BuildDialog` e sulla funzione di controllo delle finestre di dialogo.

Figura 47-17
Finestra di dialogo Apri file dati creata con lo script

```
Sub Main
   Call BuildDialog
End Sub

 'define dialog box
Sub BuildDialog
   Begin Dialog UserDialog 580,70,"Open Data File",.DialogFunc
      Text 40,7,280,21,"Data file to open:",.txtDialogTitle
      TextBox 40,28,340,21,.txtFilename
      OKButton 470,7,100,21,.cmdOK
      CancelButton 470,35,100,21,.cmdCancel
   End Dialog
   Dim dlg As UserDialog
   Dialog dlg
End Sub

  'define function that determines behavior of dialog box
Function DialogFunc(strDlgItem As String, intAction As Integer, intSuppValue As Integer) As Boolean
   Select Case intAction
      Case 1' beep when dialog is initialized
         Beep
      Case 2' value changing or button pressed
         Select Case strDlgItem
         Case "cmdOK"'if user clicks OK, open data file with specified filename
         strFilename = DlgText("txtFilename")
         Call OpenDataFile(strFilename)
         DialogFunc = False
         Case "cmdCancel"'If user clicks Cancel, close dialog
         DialogFunc = False
         End Select
   End Select
End Function
```

```
Sub OpenDataFile(strFilename As Variant)'Open data file with specified filename
  Dim objDataDoc As ISpssDataDoc
  Set objDataDoc = objSpssApp.OpenDataDoc(strFilename)
End Sub
```

Gli esempi sono inoltre disponibili nell'Aiuto in linea. Gli esempi possono essere eseguiti copiando il codice dall'Aiuto e incollandolo nella finestra dello script.

Debug degli script

I comandi del menu Debug consentono di analizzare il codice eseguendo una riga o una procedura alla volta e visualizzando il risultato. È inoltre possibile inserire punti di interruzione nelle righe dello script in corrispondenza delle quali si desidera interrompere l'esecuzione.

Per eseguire il debug di uno script automatico, aprire il corrispondente file in una finestra dello script, inserire punti di interruzione nella procedura che si desidera controllare e quindi eseguire la procedura statistica che attiva lo script automatico.

Esegui istruzione. Consente di eseguire la riga corrente. Se questa corrisponde a una chiamata di procedura o di funzione, l'esecuzione viene interrotta sulla prima riga della procedura o della funzione.

Esegui istruzione/routine. Consente di eseguire il codice fino alla riga successiva. Se questa corrisponde a una chiamata di procedura o di funzione, la procedura o la funzione viene eseguita in modo completo.

Esci da istruzione/routine. Consente di uscire dalla chiamata di procedura o di funzione corrente.

Esegui fino al cursore. Consente di eseguire il codice fino alla riga corrente.

Imposta/rimuovi interruzione. Consente di inserire o rimuovere un punto di interruzione. L'esecuzione dello script viene interrotta in corrispondenza del punto di interruzione e viene visualizzato il riquadro di debug.

Controllo immediato. Consente di visualizzare il valore dell'espressione corrente.

Aggiungi espressione di controllo. Consente di aggiungere l'espressione corrente nella finestra di controllo.

Visualizzatore oggetti. Consente di aprire il Visualizzatore oggetti.

Imposta istruzione successiva. Consente di impostare l'istruzione successiva da eseguire. È possibile selezionare solo le istruzioni incluse nella procedura/funzione corrente.

Mostra istruzione successiva. Consente di visualizzare l'istruzione successiva da eseguire.

Per analizzare uno script

- Scegliere Esegui istruzione o Esegui istruzione/routine dal menu Debug per eseguire il codice, una riga o una procedura alla volta.

  Nella finestra dello script verranno visualizzate le schede Immediate, Watch, Stack e Loaded insieme alla barra degli strumenti per il debug.

- Tramite i pulsanti della barra degli strumenti o i tasti di scelta rapida, continuare l'analisi dello script.

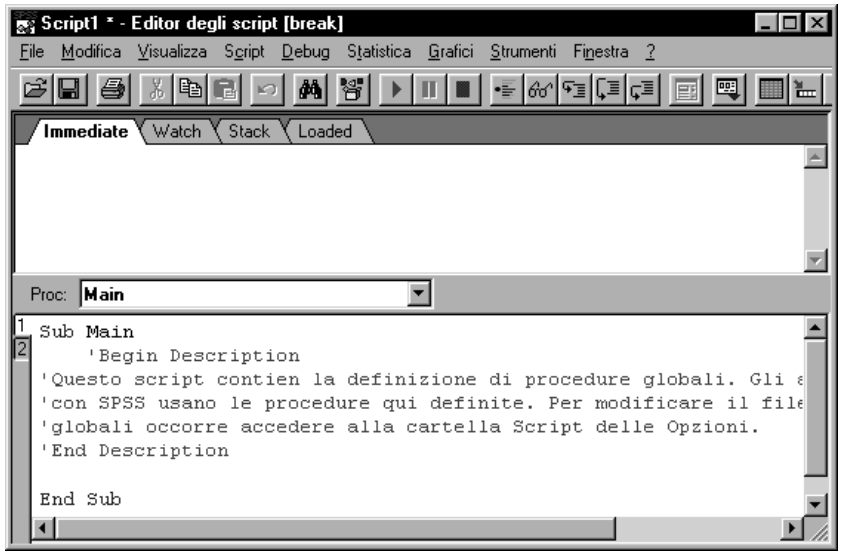
- In alternativa, è possibile selezionare Imposta/rimuovi interruzione per inserire un punto di interruzione nella riga corrente.

  L'esecuzione dello script verrà interrotta in corrispondenza del punto di interruzione.

Riquadro di debug (script)

Durante l'analisi del codice, vengono visualizzate le schede Immediate, Watch, Stack e Loaded.

Figura 47-18
Riquadro di debug nella finestra dello script


Script1 * - Editor degli script [break]
File Modifica Visualizza Script Debug Statistica Grafici Strumenti Finestra ?
Immediate
Watch
Stack
Loaded
Proc: Main
Sub Main
'Begin Description
'Questo script contien la definizione di procedure globali. Gli
'con SPSS usano le procedure qui definite. Per modificare il file
'globali occorre accedere alla cartella Script delle Opzioni.
'End Description
End Sub

Scheda Immediate. Fare clic sul nome di una variabile e quindi sul pulsante della lente d'ingrandimento per visualizzare il valore corrente della variabile. È inoltre possibile valutare un'espressione, assegnare una variabile ed eseguire una chiamata a una procedura.

- Digitare ?expr e premere Invio per visualizzare il valore di *expr*.
- Digitare var = expr e premere Invio per modificare il valore di *var*.
- Digitare subname args e premere Invio per chiamare una procedura o un'istruzione di sistema.
- Digitare Trace e premere Invio per attivare la modalità di analisi. In modalità di analisi, ogni istruzione viene stampata nella finestra Immediate durante l'esecuzione di uno script.

Scheda Watch. Per visualizzare una variabile, una funzione o un'espressione, selezionarla e scegliere Aggiungi espressione di controllo dal menu Debug. I valori visualizzati vengono aggiornati ad ogni interruzione dell'esecuzione. È possibile modificare l'espressione a sinistra di ->. Premere Invio per aggiornare immediatamente tutti i valori. Premere Ctrl+Y per eliminare la riga.

Scheda Stack. Consente di visualizzare le righe in cui è stata richiamata l'istruzione corrente. La prima riga corrisponde all'istruzione corrente, la seconda alla riga da cui è stata richiamata la prima e così via. Facendo clic su una riga, è possibile visualizzarla nella finestra di modifica.

Scheda Loaded. Consente di visualizzare un elenco degli script attivi. Facendo clic su una riga, è possibile visualizzare lo script corrispondente.

File di script e di sintassi

I file di sintassi, con estensione *sps*, non sono uguali ai file di script, con estensione *sbs*. I file di sintassi includono comandi scritti nel linguaggio a comandi SPSS, che consente di eseguire procedure statistiche di SPSS e trasformazioni di dati. Mentre gli script consentono di gestire l'output e di automatizzare altre operazioni in genere eseguite usando l'interfaccia grafica di menu e finestre di dialogo, il linguaggio a comandi SPSS fornisce un metodo alternativo per comunicare direttamente con il backend SPSS, la parte del sistema che gestisce i calcoli statistici e le trasformazioni dei dati.

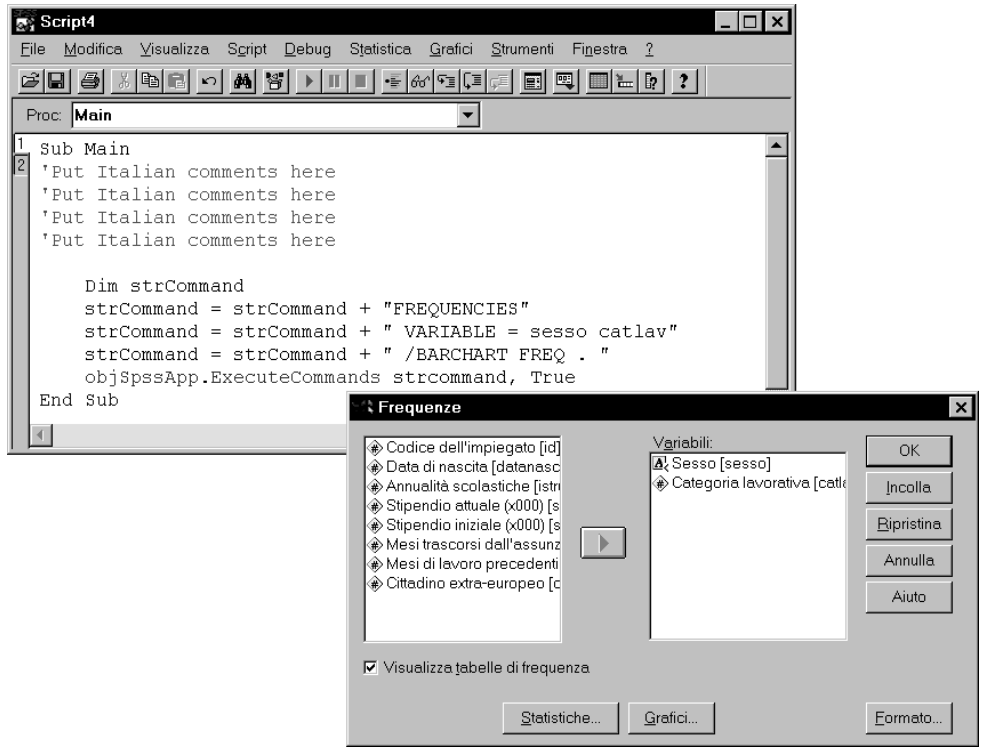
È possibile combinare file di script e di sintassi per ottenere un livello di flessibilità ancora maggiore eseguendo uno script dalla sintassi dei comandi di SPSS o incorporando tale sintassi in uno script.

Esecuzione della sintassi dei comandi da uno script

È possibile eseguire la sintassi dei comandi di SPSS da uno script di automazione tramite il metodo `ExecuteCommands`. La sintassi dei comandi di SPSS consente di creare grafici ed eseguire trasformazioni di dati e procedure statistiche. La maggior parte di queste funzioni non può essere automatizzata direttamente in script.

Il metodo più semplice per creare un file di sintassi dei comandi consiste nell'eseguire selezioni nelle finestre di dialogo di SPSS e quindi incollare la sintassi per le selezioni nella finestra dello script.

Figura 47-19
Sintassi dei comandi incollata in uno script



Quando si accede alle finestre di dialogo dai menu della finestra dello script, il pulsante Incolla inserisce tutto il codice necessario per eseguire i comandi all'interno dello script.

Nota: Per aprire la finestra di dialogo è necessario utilizzare i menu della finestra dello script. In caso contrario, i comandi vengono incollati in una finestra di sintassi anziché nella finestra dello script.

Per incollare la sintassi dei comandi SPSS in uno script

- Nella finestra dello script, scegliere i comandi dei menu Statistiche, Grafici e Strumenti per aprire le finestre di dialogo.
- Nella finestra di dialogo eseguire le selezioni desiderate.

▶ Fare clic su Incolla.

Nota: Per aprire la finestra di dialogo è necessario utilizzare i menu della finestra dello script. In caso contrario, i comandi vengono incollati in una finestra di sintassi anziché nella finestra dello script.

Esecuzione di uno script dalla sintassi dei comandi

Il comando `SCRIPT` consente di eseguire uno script all'interno della sintassi dei comandi. Specificare il nome dello script che si desidera eseguire, racchiudendo il nome del file tra virgolette, come indicato di seguito:

```
SCRIPT 'C:\PROGRAM FILES\SPSS\CLEAN NAVIGATOR.SBS'.
```

Per informazioni dettagliate sulla sintassi, vedere *SPSS Command Syntax Reference*.

Capitolo
48
Sistema di gestione dell'output

Sistema di gestione dell'output (SGO) consente di scrivere automaticamente le categorie di output selezionate in file di output in formati diversi, che includono:

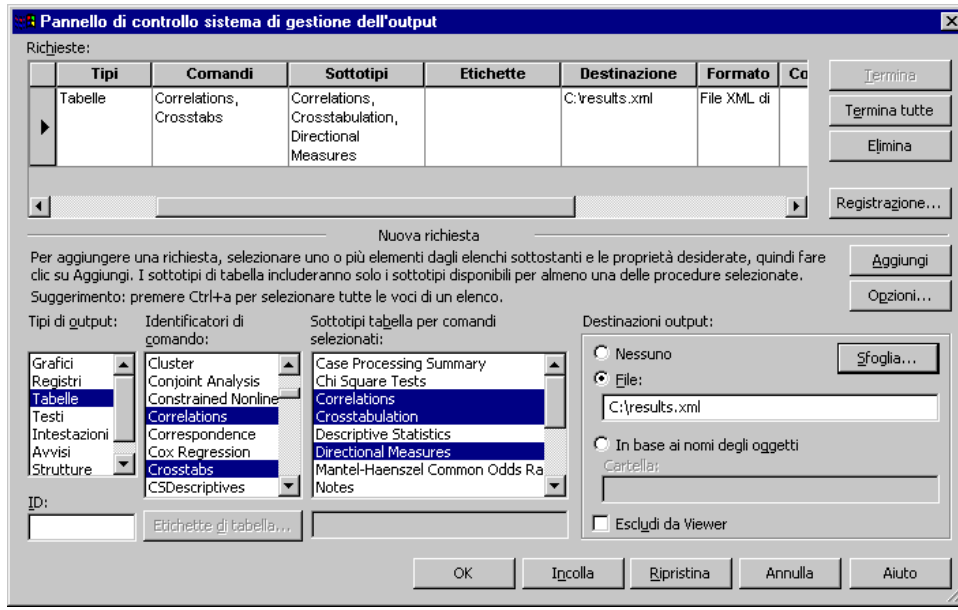
- **Formato file dati di SPSS (sav).** L'output visualizzato in tabelle pivot nel Viewer può essere scritto nel formato di un file dati di SPSS e può essere utilizzato come input per i comandi successivi.
- **XML.** Le tabelle, l'output testuale e molti grafici possono essere scritti in formato XML.
- **HTML**. Le tabelle e l'output testuale possono essere scritti in formato HTML. I grafici standard (non interattivi) ed i diagrammi di modelli di alberi (opzione Albero decisionale) possono essere inclusi come file d'immagine. I file d'immagine vengono salvati in una sottodirectory a parte (cartella).
- **Testo.** Le tabelle e l'output testuale possono essere scritti come testo delimitato da tabulazioni o da spazi.

Per utilizzare il Pannello di controllo del sistema di gestione dell'output

▶ Dai menu, scegliere:

Strumenti
  Panello di controllo di SGO...

Figura 48-1
Pannello di controllo sistema di gestione dell'output



È possibile usare il pannello di controllo per avviare o arrestare il trasferimento dell'output a varie destinazioni.

- Tutte le richieste SGO rimangono attive fino quando non vengono terminate o fino alla fine della sessione.
- Il file di destinazione specificato in una richiesta SGO viene reso disponibile alle altre procedure SPSS ed alle altre applicazioni solo al termine della richiesta SGO.
- Poiché i file di destinazione specificati vengono salvati nella memoria (RAM) quando la richiesta è attiva, è possibile che le richieste SGOattive che scrivono grandi volumi di output sui file esterni utilizzino molta memoria.
- Le richieste SGO sono indipendenti le une dalle altre. Lo stesso output può essere trasmesso a posizioni diverse in formati diversi, a seconda delle specifiche delle singole richieste SGO.

- L'ordine degli oggetti dell'output di una destinazione specifica varia a seconda dell'ordine e del funzionamento delle procedure che generano l'output.
- SGO non è in grado di trasmettere grafici o oggetti Warning creati da procedure grafiche interattive (menu Grafici, sottomenu Interattivi) oppure mappe create con le procedure di associazione (menu Grafici, sottomenu Mappa).

Per aggiungere nuove richieste SGO

▶ Selezionare i tipi di output (tabelle, grafici, ecc.) da includere. (Per ulteriori informazioni, vedere "Tipi di oggetto di output" a pag. 715.)

▶ Selezionare i comandi da includere. Per includere tutto l'output, selezionare tutte le voci dell'elenco. (Per ulteriori informazioni, vedere "Identificatori di comando e sottotipi di tabelle" a pag. 717.)

▶ Per i comandi che producono una tabella pivot come output, selezionare i tipi di tabelle specifici da includere.

L'elenco visualizza solo le tabelle disponibili nei comandi selezionati. Quindi, l'elenco visualizza tutti ti tipi di tabelle disponibili in uno o più dei comandi selezionati. Se non sono stati selezionati comandi, vengono visualizzati tutti i tipi di tabelle. (Per ulteriori informazioni, vedere "Identificatori di comando e sottotipi di tabelle" a pag. 717.)

▶ Per selezionare le tabelle in base alle etichette del testo anziché in base ai sottotipi, fare clic su Etichette. (Per ulteriori informazioni, vedere "Etichette" a pag. 718.)

▶ Fare clic su Opzioni per specificare il formato dell'output (ad esempio file dati SPSS, XML, HTML). Per impostazione predefinita, viene usato il formato File XML di output. Per ulteriori informazioni, vedere "Opzioni SGO" a pag. 720.)

▶ Specificare la destinazione di output:

- **File.** Tutto l'output selezionato viene indirizzato a un unico file.

- **Basati su nomi oggetto.** L'output viene inviato a più file di destinazione in base ai nomi degli oggetti. Viene creato un file diverso per ciascun oggetto di output, con un nome di file basato o sui nomi del sottotipo di tabella o sulle etichette delle tabelle. Immettere il nome della cartella di destinazione.
- **Nuovo file di dati.** L'output in formato di file di dati SPSS può essere inviato a un file di dati. Il file di dati può quindi essere usato nell'ambito della stessa sessione, ma non viene salvato a meno che non si decida esplicitamente di salvarlo al termine della sessione. Questa opzione è disponibile solo per l'output in formato di file di dati SPSS. I nomi dei file di dati devono essere conformi alle regole dei nomi delle variabili di SPSS. Per ulteriori informazioni, vedere "Nomi delle variabili" in Capitolo 5 a pag. 99.

▶ Oppure:

- Escludere l'output selezionato da Viewer. Se si seleziona Escludi da Viewer, i tipi di output della richiesta SGO non vengono visualizzati nella finestra Viewer. Se più richieste SGO attive includono gli stessi tipi di output, i tipi di output visualizzati in Viewer variano a seconda della richiesta SGO più recente che contiene i tipi di output specificati. Per ulteriori informazioni, vedere "Esclusione della visualizzazione dell'output in Viewer" a pag. 726.
- Assegnare una stringa ID alla richiesta. A tutte le richieste viene automaticamente assegnato un valore ID, sebbene sia sempre possibile sostituire la stringa ID predefinita del sistema con l'ID descrittivo. Questa operazione può essere utile per identificare più facilmente più richieste attive. I valori ID assegnabili non possono iniziare con il segno Dollaro ($).

Di seguito vengono riportati una serie di consigli utili per selezionare più voci negli elenchi:

- Premere Ctrl-A per selezionare tutte le voci dell'elenco.
- Premere Maiusc. ed il pulsante del mouse per selezionare più voci contigue.
- Premere Ctrl ed il pulsante del mouse per selezionare più voci non contigue.

Per terminare ed eliminare richieste SGO

Le nuove richieste SGO attive vengono visualizzate nell'elenco Richieste, a partire dalla richiesta più recente. Per modificare la larghezza delle colonne contenenti informazioni, è sufficiente selezionare e trascinare i bordi, quindi scorrere

orizzontalmente l'elenco per visualizzare ulteriori informazioni sulla richiesta desiderata.

L'asterisco (*) dopo la parola Attivo nella colonna *Stato* indica che la richiesta SGO è stata creata con una sintassi di comando che comprende funzionalità non disponibili nel Pannello di comando.

Per terminare una richiesta SGO attiva specifica:

- Nell'elenco Richieste fare clic su qualsiasi cella della riga relativa alla richiesta.
- Fare clic su Termina.

Per terminare tutte le richieste SGO attive:

- Fare clic su Termina tutte.

Per eliminare una nuova richiesta (ossia una richiesta aggiunta ma non ancora attiva):

- Nell'elenco Richieste fare clic su qualsiasi cella della riga relativa alla richiesta.
- Fare clic su Elimina.

Nota: le richieste SGO attive vengono terminate solo quando si seleziona OK.

Tipi di oggetto di output

Esistono sette diversi tipi di oggetti di output:

Grafici. Grafici, ad eccezione dei grafici interattivi e delle mappe. Gli oggetti grafico vengono inclusi solo nei formati di destinazione XML e HTML. Per il formato HTML, i file d'immagine dei grafici vengono salvati in una sottodirectory a parte (cartella).

Registri. Oggetti Testo registri. Questi oggetti contengono alcuni tipi di errori e messaggi di avviso. A seconda delle impostazioni selezionate in Opzioni (menu Modifica, Opzioni, scheda Viewer), questi oggetti possono contenere anche la sintassi dei comandi eseguita durante la sessione. Gli oggetti registro sono etichettati *Registro* nella vista riassuntiva di Viewer.

Tabelle. Le tabelle pivot in Viewer (con le tabelle Note). Le tabelle sono gli unici oggetti di output che possono essere trasmessi in formato file dati SPSS (*.sav*).

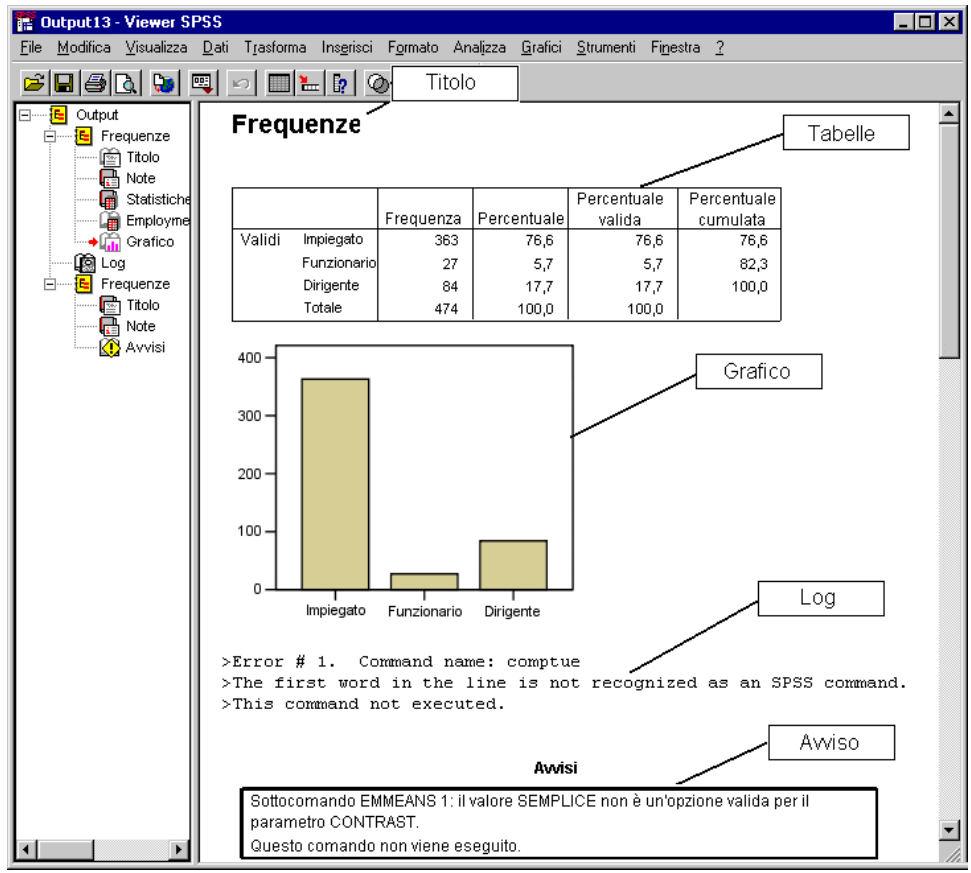
Testo. Oggetti di testo diversi da ' registri o titoli (compresi oggetti etichettati come *Output di testo* nella vista riassuntiva del Viewer.

Alberi. Diagrammi dei modelli di alberi creati tramite l'opzione Albero decisionale. Gli oggetti albero sono inclusi solo nei formati di destinazione XML e HTML.

Intestazioni. Oggetti di testo etichettati come *titolo* nella vista riassuntiva del Viewer. Gli oggetti Intestazione non vengono inclusi nel formato di output XML.

Avvisi. Oggetti Avviso. Questi oggetti contengono alcuni tipi di errore e messaggi di avviso.

Figura 48-2
Tipi di oggetti di output



Identificatori di comando e sottotipi di tabelle

Identificatori di comando

Gli identificatori di comando sono disponibili in tutte le procedure per la creazione di grafici e statistiche nonché in qualsiasi altro comando che blocchi l'output con un'intestazione identificabile nella vista riassuntiva di Viewer. Questi identificatori sono generalmente, ma non sempre, equivalenti o simili ai nomi delle procedure dei menu o nei titoli delle finestre di dialogo, che sono generalmente, ma non sempre, simili ai nomi dei comandi SPSS associati. L'identificatore del comando per la procedura Frequenze è "Frequenze", che corrisponde anche al nome del comando associato.

Tuttavia, in alcuni casi il nome della procedura non corrisponde a quello dell'identificatore del comando e/o al nome del comando. Esempio: tutte le procedure del sottomenu Test Nonparametrici (nel menu Analizza) utilizzano lo stesso comando di base, ma l'identificatore del comando è uguale al nome del comando di base, Test non parametrici.

Sottotipi di tabelle

I sottotipi di tabelle rappresentano i diversi tipi di tabelle pivot che possono essere prodotti. Alcuni sottotipi possono essere prodotti con un solo tipo di comando, mentre altri possono essere prodotti con più tipi di comandi, sebbene le tabelle abbiano lo stesso aspetto. Sebbene i nomi dei sottotipi di tabelle siano descrittivi, sono generalmente disponibili varie opzioni (soprattutto se si è selezionato un numero elevato di comandi); quindi è possibile che due sottotipi abbiano nomi simili.

Per ricercare identificatori di comandi e sottotipi di tabelle

In caso di dubbio, è possibile ricercare i nomi degli identificatori di comando e dei sottotipi delle tabelle nella finestra di Viewer.

- Eseguire la procedura per generare l'output in Viewer.
- Fare clic con il pulsante destro del mouse nella vista riassuntiva di Viewer.
- Selezionare Copia identificatore comandi SGO oppure Copia sottotipo tabelle SGO.

- Incollare l'identificatore di comando o il nome del sottotipo di tabelle in un editor di testo (ad esempio in una finestra di sintassi SPSS).

Etichette

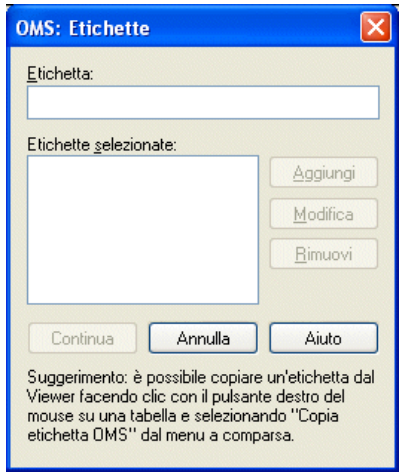
Anziché specificare i nomi dei sottotipi di tabelle, è possibile selezionare le tabelle in base al testo visualizzato nella vista riassuntiva di Viewer. È possibile anche selezionare altri tipi di oggetto in base alle relative etichette. Le etichette consentono di distinguere tra più tabelle dello stesso tipo il cui testo riassuntivo rifletta alcuni attributi dell'oggetto di output; ad esempio i nomi o le etichette delle variabili. Il testo delle etichette può tuttavia essere influenzato da numerosi fattori:

- Se è attiva la distinzione per gruppi, è possibile che all'etichetta venga aggiunta l'identificazione del gruppo di file distinto in base alla variabile categoriale.
- Le etichette che contengono informazioni sulle variabili o i valori sono influenzate dalle impostazioni delle opzioni selezionate per le etichette di output (menu Modifica, Opzioni, Etichette di output).
- Le etichette sono influenzate dall'impostazione corrente della lingua di output (menu Modifica, Opzioni, scheda Generale).

Per specificare le etichette da usare per identificare gli oggetti di output

- Nel Pannello di controllo di SGO, selezionare uno o più tipi di output, quindi uno o più tipi di comandi.
- Fare clic su Etichette.

Figura 48-3
Finestra di dialogo Etichette di SGO



- Digitare l'etichetta *esattamente* come appare nella vista riassuntiva della finestra Viewer. In alternativa, è possibile anche fare clic con il pulsante destro del mouse sulla voce nella vista riassuntiva, selezionare Copia etichetta SGO ed incollare l'etichetta copiata nel campo di testo Etichetta.
- Fare clic su Aggiungi.
- Ripetere l'operazione per ciascuna etichetta che si desidera includere.
- Fare clic su Continua.

Caratteri jolly

È possibile utilizzare un asterisco (*) per l'ultimo carattere della stringa dell'etichetta in modo che funga da carattere jolly. In questo modo vengono selezionate tutte le etichette che iniziano con la stringa specificata (ad eccezione dell'asterisco). Ciò vale solo se l'asterisco è l'ultimo carattere, poiché gli asterischi possono essere utilizzati come caratteri validi per le etichette.

Opzioni SGO

È possibile utilizzare la finestra di dialogo Opzioni SGO per:

- Specificare il formato dell'output.
- Includere o escludere l'output di un diagramma di modello di albero e specificarne il formato grafico.
- Specificare quali elementi della dimensione della tabella devono essere inclusi nella dimensione della riga.
- Per il formato file dati SPSS, includere una variabile che identifichi il numero della tabella sequenziale che rappresenta l'origine di ciascun caso.

Per specificare le opzioni SGO:

▶ Fare clic su Opzioni nel Pannello di controllo sistema di gestione dell'output.

Figura 48-4
Finestra di dialogo Opzioni SGO

Formato

File XML di output. File XML conforme allo schema *Output SPSS*. I grafici standard vengono inclusi come file XML conforme allo schema*vizml*. È possibile anche esportare tutti i grafici e le mappe come file singoli nel formato grafico selezionato.

HTML. Gli oggetti che vengono rappresentati come tabelle pivot in Viewer vengono convertiti in semplici tabelle HTML. Gli attributi TableLook (caratteristiche dei caratteri, stili dei bordi, colori, ecc.) non sono supportati. Agli oggetti di output di test viene assegnato il tag <PRE> nel file HTML. Se si decide di includere i grafici, questi vengono esportati come singoli file nel formato grafico selezionato ed incorporati per riferimento (<IMG SRC = 'nomefile.est'>) nel documento HTML. I file d'immagine vengono salvati in una sottodirectory a parte (cartella).

File dati SPSS. File in formato binario. Tutti i tipi di oggetti di output diversi dalle tabelle sono esclusi. Ciascuna colonna della tabella diventa una variabile nel file dati. Per utilizzare un file dati creato con SGO nella stessa sessione, è necessario terminare la richiesta SGO attiva prima di aprire il file dati. Per ulteriori informazioni, vedere "Trasmissione dell'output ai file dati SPSS" a pag. 726.

SVWS XML. XML utilizzato da SmartViewer Web Server. File JAR/ZIP che contiene file XML, CSV e di altro formato. SmartViewer Web Server è un prodotto a parte basato su server.

Testo. Testo separato da spazi. L'output viene scritto come testo con l'output della tabella allineato agli spazi per i caratteri con passo fisso. Sono esclusi tutti i grafici e le tabelle.

Testo con tabulazioni. Testo delimitato da tabulazioni. Le tabulazioni delimitano gli elementi delle colonne delle tabelle per l'output che verrebbe visualizzato come tabelle pivot in Viewer. Le righe dei blocchi di testo vengono scritte come tali senza che vengano inserite tabulazioni per dividerle nei punti più opportuni. Sono esclusi tutti i grafici e le tabelle.

Immagini grafiche

Nel formato HTML, è possibile includere grafici (esclusi quelli interattivi) e diagrammi di modelli di albero come file di immagine. Per ciascun grafico e/o albero viene creato un file d'immagine a parte ed in ciascun file d'immagine vengono inclusi i tag standard <IMG SRC='nomefile'>.

I file d'immagine vengono salvati in una sottodirectory a parte (cartella). Il nome della sottodirectory corrisponde al nome del file di destinazione HTML senza estensioni, al quale viene aggiunto *_files*. Ad esempio, se il file di destinazione HTML è *julydata.htm*, la sottodirectory delle immagini verrà chiamata *julydata_files*.

- **Formato.** I file di immagini disponibili sono PNG, JPG, EMF e BMP.
- **Dimensioni.** È possibile scalare le dimensioni dell'immagine nell'intervallo 10% - 200%.

Tabella pivot

Per l'output in formato tabella pivot, è possibile specificare gli elementi delle dimensioni che devono essere visualizzati nelle colonne. Tutte le altre dimensioni vengono visualizzate sulle righe. Nel formato file dati SPSS, le colonne delle tabelle diventano variabili, mentre le righe diventano casi.

- Se si specificano elementi con dimensioni diverse per le colonne, questi vengono annidati nell'ordine in cui sono presenti nell'elenco. Per il formato file dati SPSS, i nomi delle variabili vengono creati utilizzano gli elementi delle colonne annidate. Per ulteriori informazioni, vedere "Nomi di variabili nei file dati generati da SGO" a pag. 735.
- Se la tabella non 'contiene elementi relativi alle dimensioni, questi vengono visualizzati sulle righe.
- I pivot tabella specificati in questo contesto non hanno alcun effetto sulle tabelle visualizzate in Viewer.

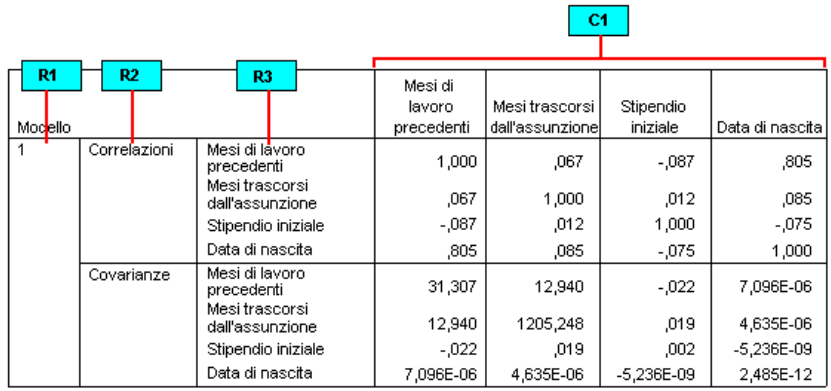
Ciascuna dimensione di una tabella - riga, colonna, strato - possono contenere più elementi o nessun elemento. Ad esempio una tavola di contingenza contiene una dimensione in un'unica riga e una dimensione in una sola colonna; ciascuna di queste dimensioni contiene una delle variabili usate nella tabella. È possibile usare un argomento relativo alla posizione o "nomi" di dimensioni per specificare le dimensioni che si desidera inserire nella dimensione della colonna.

Tutte le dimensioni nelle righe. Crea una singola riga per ciascuna tabella. Per i file dati in formato SPSS, ciò significa che ciascuna tabella rappresenta un caso singolo e che tutti gli elementi della tabella sono variabili.

Elenco posizioni. Un argomento relativo alla posizione è generalmente rappresentato da una lettera che indica la posizione predefinita dell'elemento - C per la colonna, R per la riga o L per lo strato - seguita da un numero intero positivo che indica la posizione predefinita della dimensione. R1, ad esempio, indica la dimensione della riga più esterna.

- Per specificare più dimensioni, è necessario separare ciascuna dimensione con uno spazio - ad esempio R1 C2.
- Se la lettera della dimensione è seguita da ALL, significa che tutti gli elementi della dimensione sono disposti nell'ordine predefinito. CALL, ad esempio, è uguale al comportamento predefinito poiché utilizza tutti gli elementi della colonna nell'ordine predefinito per creare le colonne.
- CALL RALL LALL (o RALL CALL LALL, etc.) inserisce tutte le dimensioni nelle colonne. Nel formato file dati SPSS, ciò crea una riga/un caso per tabella nel file dati.

Figura 48-5
Argomenti relativi alla posizione di righe e colonne



Modello			Mesi di lavoro precedenti	Mesi trascorsi dall'assunzione	Stipendio iniziale	Data di nascita
1	Correlazioni	Mesi di lavoro precedenti	1,000	,067	-,087	,805
		Mesi trascorsi dall'assunzione	,067	1,000	,012	,085
		Stipendio iniziale	-,087	,012	1,000	-,075
		Data di nascita	,805	,085	-,075	1,000
	Covarianze	Mesi di lavoro precedenti	31,307	12,940	-,022	7,096E-06
		Mesi trascorsi dall'assunzione	12,940	1205,248	,019	4,635E-06
		Stipendio iniziale	-,022	,019	,002	-5,236E-09
		Data di nascita	7,096E-06	4,635E-06	-5,236E-09	2,485E-12

Elenco nomi di dimensione. In alternativa agli argomenti relativi alla posizione, è possibile utilizzare i "nomi" delle dimensioni rappresentati dalle etichette di testo visualizzate nella tabella. Ad un esempio una tavola di convergenza semplice bidimensionale contiene una dimensione di una singola riga e una dimensione di una singola colonna, entrambe con etichette basate sulle variabili contenute nelle dimensioni specificate, oltre ad una dimensione di uno strato singolo etichettato *Statistiche* (se è stato impostato l'italiano come lingua di output).

- I nomi delle dimensioni possono variare a seconda della lingua di output e/o delle impostazioni che influiscono sulla visualizzazione dei nomi delle variabili e/o delle etichette nelle tabelle.
- Ciascun nome di dimensione deve essere racchiuso tra virgolette semplici o doppie. Per specificare nomi di più dimensioni, è necessario aggiungere uno spazio tra ciascun nome.

Le etichette associate alle dimensioni non sono sempre ovvie.

Per visualizzare tutti gli elementi dimensione e le relative etichette di una tabella pivot

▶ Attivare (con un doppio clic) del mouse la tabella in Viewer.

▶ Dai menu, scegliere:

Visualizza
  Mostra tutto

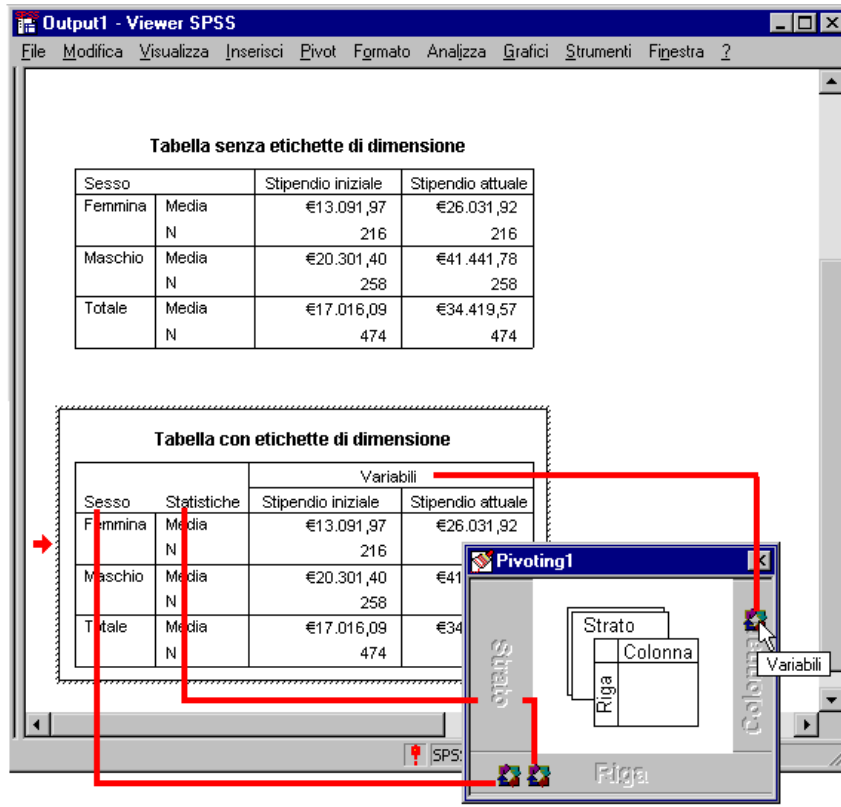
e/o

▶ Se Pivoting ' non è visualizzato, dai menu scegliere:

Pivot
  Pivoting

▶ Posizionare il puntatore del mouse sui pivoting per visualizzare l'etichetta.

Figura 48-6
Nomi delle dimensioni visualizzate nella tabella e nei pivot



Registrazione

È possibile registrare un'attività SGO in un registro in formato XML o testo.

- Il registro tiene traccia di tutte le nuove richieste SGO della sessione, ma non include le richieste SGO che erano già attive prima della richiesta di un registro.
- Il file di registro corrente termina se si specifica un nuovo file di registro o se si deseleziona l'opzione Registra attività SGO.

Per specificare una registrazione SGO:

► Fare clic su Registrazione nel Pannello di controllo sistema di gestione dell'output.

Esclusione della visualizzazione dell'output in Viewer

La casella di controllo Escludi da Viewer ha effetto su tutto l'output selezionato nella richiesta SGO, poiché evita che tale output venga visualizzato nella finestra del Viewer. Questo processo è spesso utile per i lavori di produzione che generano molti output e per i quali non si desidera visualizzare i risultati in formato di documento Viewer (file *.spo*). L'opzione può anche essere usata per nascondere oggetti di output specifici che non si desiderano visualizzare senza dover trasmettere l'output ad un file e formato esterno.

Per nascondere oggetti di output specifici senza trasmettere altri output ad un file esterno:

- Creare una richiesta SGO che identifichi l'output indesiderato.
- Selezionare Escludi da Viewer.
- Per la destinazione di output selezionare File, lasciando il campo File vuoto.
- Fare clic su Aggiungi.

L'output selezionato viene escluso da Viewer, mentre tutti gli altri output vengono visualizzati in Viewer nel formato standard.

Trasmissione dell'output ai file dati SPSS

I file dati SPSS sono costituiti dalle variabili delle colonne e dai casi delle righe, che rappresenta essenzialmente la modalità con cui le tabelle pivot vengono convertite in file dati:

- Le colonne della tabella sono le variabili del file dati. I nomi delle variabili validi vengono creati utilizzando le etichette delle colonne.
- Le etichette delle righe della tabella diventano i nomi di variabili generiche (*Var1*, *Var2*, *Var3*, ecc.) nel file di dati. Il valore di queste variabili è rappresentato dalle etichette delle righe nella tabella.
- Nel file dati vengono automaticamente incluse tre variabili identificatore della tabella: *Command_*, *Subtype_* e *Label_*. Tutte e tre sono variabili stringa. Le prime due variabili corrispondono al comando e agli identificatori del sottotipo.

Per ulteriori informazioni, vedere "Identificatori di comando e sottotipi di tabelle" a pag. 717. *Label_* contiene il testo del titolo della tabella.

- Le righe della tabella diventano i casi del file di dati.

Esempio: Tabella bidimensionale singola

Nel caso più semplice, rappresentato da una tabella bidimensionale singola, le colonne della tabella diventano variabili, mentre le righe diventano i casi del file di dati.

Figura 48-7
Tabella bidimensionale singola

Sesso		Stipendio attuale	Stipendio iniziale
Femmina	Media	€26.031,92	€13.091,97
	Mediana	€24.300,00	€12.375,00
	Minimo	€15.750	€9.000
	Massimo	€58.125	€30.000
Maschio	Media	€41.441,78	€20.301,40
Totale			

temp.sav - SPSS Editor dei dati
File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?
14 : Command_
Visualizzazione dati
Visualizzazione variabili
SPSS Processore pronto

	Command_	Subtype_	Label_	Var1	Var2	Stipendio attua	Stipendio iniziale
1	Means	Report	Report	Femmina	Media	€26.031,92	€13.091,97
2	Means	Report	Report	Femmina	Media	€24.300,00	€12.375,00
3	Means	Report	Report	Femmina	Minimo	€15.750,00	€9.000,00
4	Means	Report	Report	Femmina	Massimo	€58.125,00	€30.000,00
5	Means	Report	Report	Maschio	Media	€41.441,78	€20.301,40
6	Means	Report	Report	Maschio	Media	€32.850,00	€15.750,00
7	Means	Report	Report	Maschio	Minimo	€19.650,00	€9.000,00
8	Means	Report	Report	Maschio	Massimo	€135.000,00	€79.980,00
9	Means	Report	Report	Totale	Media	€34.419,57	€17.016,09
10	Means	Report	Report	Totale	Media	€28.875,00	€15.000,00
11	Means	Report	Report	Totale	Minimo	€15.750,00	€9.000,00

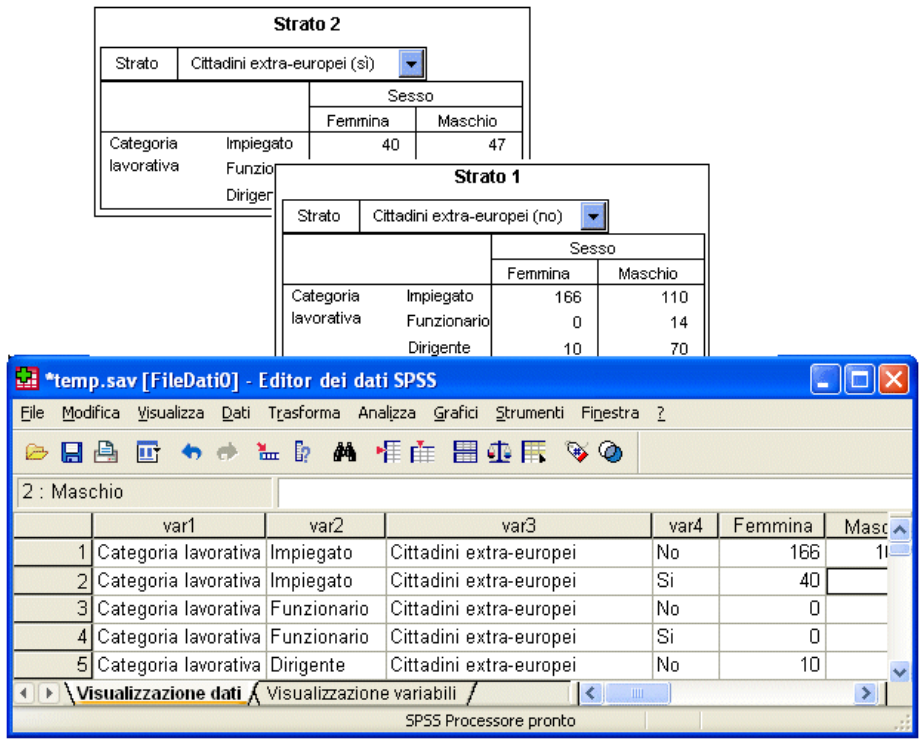
- Le prime tre variabili identificano la tabella di origine in base al sottotipo di comando e all'etichetta.

- Ai due elementi che definivano le righe nella tabella, ossia i valori della variabile *Sesso*, e le misure statistiche vengono assegnati nomi di variabili generici *Var1* and *Var2*. Questi variabili sono entrambe variabili stringa.
- Le etichette delle colonne della tabella vengono usate per creare nomi di variabili valide. In questo caso, i nomi delle variabili sono basati sulle etichette delle variabili delle tre variabili scala riassunte nella tabella. Se le variabili non hanno' un'etichetta variabile definita o si sceglie di visualizzare i nomi delle variabili anziché le etichette delle variabili come etichette delle colonne della tabella, i nomi delle variabili nel nuovo file saranno gli stessi del file dati di origine.

Esempio: Tabelle con strati

Oltre a righe e colonne una tabella può contenere anche un terzo elemento: lo strato.

Figura 48-8
Tabella con strati



- La variabile denominata *Classificazione per minoranza* definisce gli strati nella tabella. Nel file di dati vengono create altre due variabili: la prima identifica lo strato, mentre la seconda identifica le categorie degli elementi dello strato.
- Al pari delle variabili create con le righe, le variabili create con gli strati sono variabili stringa con nomi generici (formato dal prefisso *Var* seguito da un numero progressivo).

File dati creati con tabelle multiple

Se si associano più tabelle allo stesso file dati, ciascuna tabella viene aggiunta al file dati con modalità simili a quelle utilizzate per unire i file di dati aggiungendo casi da un file di dati ad un altro (menu Dati, Unisci file, Aggiungi casi).

- Ciascuna tabella successiva aggiungerà sempre i casi al file di dati.
- Se le etichette delle colonne della tabella differiscono, è possibile anche che ciascuna tabella aggiunga variabili al file di dati utilizzando anche i valori mancanti dei casi di altre tabelle che 'non hanno etichette simili per le colonne.

Esempio: tabelle multiple con le stesse etichette colonne

Le tabelle multiple che contengono le stesse etichette per le colonne forniscono generalmente i file di dati più utili (ossia file di dati che non richiedono 'ulteriori manipolazioni). Due o più tabelle della frequenza generate tramite la procedura Frequenza possono avere, ad esempio, etichette identiche per le colonne.

Figura 48-9
Due tabelle con etichette identiche per le colonne

Sesso

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Valido	Femmina	216	45.6	45,6	45,6
	Maschio	258	54.4	54,4	100,0

Categoria lavorativa

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Valido	Impiegato	363	76,6	76,6	76,6
	Funzionario	27	5,7	5,7	82,3
	Dirigente	84	17,7	17,7	100,0

*temp.sav [FileDati0] - Editor dei dati SPSS

File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?

1 : Command_ | Frequenza

	Subtype_	Label_	Var1	Var2	Frequenza	Percentuale	Percentuale_val
1	Frequenza	Sesso	Valido	Femmina	216	45,6	45,6
2	Frequenza	Sesso	Valido	Maschio	258	54,4	54,4
3	Frequenza	Sesso	Valido	Totale	474	100,0	100,0
4	Frequenza	Categoria lavorativa	Valido	Impiegato	363	76,6	76,6
5	Frequenza	Categoria lavorativa	Valido	Funzionario	27	5,7	5,7

Visualizzazione dati | Visualizzazione variabili

SPSS Processore pronto

- La seconda tabella fornisce ulteriori casi (righe) al file di dati ma nessuna variabile nuova perché le etichette delle colonne sono identiche, quindi grosse sezioni di dati mancanti.
- Sebbene i valori di *Command_* e *Subtype_* siano identici, il valore di *Label_* identifica la tabella di origine per ciascun grupo di casi perché le due tabelle relative alla frequenza hanno titoli diversi.

Esempio: tabelle multiple con etichette di colonna diverse

Nel file di dati viene creata una nuova variabile per ciascuna etichetta di colonna univoca delle tabelle associate al file di dati. Questo processo genera blocchi di valori mancanti se le tabelle contengono etichette colonne diverse.

Figura 48-10
Due tabelle con etichette di colonna diverse

Tabella 1

Sesso		Stipendio iniziale	Stipendio attuale
Femmina	Media	€13.091,97	€26.031,92
	Mediano	€12.375,00	€24.300,00
Maschio	Media	€20.301,40	€41.441,78
Totale			

Tabella 2

Sesso		Livello di istruzione	Mesi da noleggio
Femmina	Media	12,37	80,38
	Mediano	12,00	81,00
Maschio	Media	14,43	81,72

Senza titolo1 [FileDati0] - Editor dei dati SPSS

File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?

1 :

	Subtype_	Label_	Var1	Var2	Stipendi attuale	Stipendio iniziale	Livello di istruzione	Mesi da noleggio
1	Report	Report	Femmina	Mean	€26032	€13092,0	.	.
2	Report	Report	Femmina	Median	€24300	€12375,0	.	.
3	Report	Report	Maschio	Media	€41442	€20301,4	.	.
4	Report	Report	Maschio	Mediano	€32850	€15750,0	.	.
5	Report	Report	Totale	Media	€34420	€17016,1	.	.
6	Report	Report	Totale	Median	€28875	€15000,0	.	.
7	Report	Report	Femmina	Media	.	.	12,37	80,38
8	Report	Report	Femmina	Mediano	.	.	12,00	81,00
9	Report	Report	Maschio	Media	.	.	14,43	81,72
10	Report	Report	Maschio	Mediano	.	.	15,00	82,00
11	Report	Report	Totale	Media	.	.	13,49	81,11
12	Report	Report	Totale	Mediano	.	.	12,00	81,00

Visualizzazione dati | Visualizzazione variabili

- La prima tabella ha due colonne denominate rispettivamente *Stipendio iniziale* e *Stipendio attuale* che non sono presenti nella seconda tabella e che generano valori mancanti per le variabili dei casi della seconda tabella.
- La seconda tabella ha due colonne denominate rispettivamente *Livello di educazione* e *Mesi di servizio* che non sono presenti nella prima tabella e che provocano valori mancanti per le variabili dei casi della prima tabella.
- Differenze nella variabili simile a quelle dell'esempio possono verificarsi anche per tabelle dello stesso sottotipo. Le tabelle dell'esempio sono infatti entrambe dello stesso sottotipo.

Esempio: file di dati non creati con tabelle multiple

I file di dati non vengono creati per le tabelle che non hanno lo stesso numero di elementi riga. Il numero di righe non ' deve essere necessariamente lo stesso, ma il numero di *elementi* riga che si trasformano nel file di dati deve sempre essere uguale. Esempio: una tabella incrociata con due variabili e una tabella incrociata con tre variabili contengono numeri di elementi riga diversi, poiché la variabile "strato" è nidificata nella variabile riga della vista della tabella predefinita con tre variabili.

Figura 48-11
Tabelle con numero di elementi riga diversi

Categoria lavorativa * Tabella di contingenza Genere

		Sesso		
		Femmina	Maschio	Totale
Categoria lavorativa	Impiegato	206	157	363
	Funzionario	0	27	27
	Dirigente	10	74	84
Totale		216	258	474

Categoria lavorativa * Genere * Tabella di contingenza Cittadini extra-europei

			Sesso		
Cittadini extra-europei			Femmina	Maschio	Totale
No	Categoria lavorativa	Impiegato	166	110	276
		Funzionario	0	14	14
		Dirigente	10	70	80
	Totale		176	194	370
Sì	Categoria lavorativa	Impiegato	40	47	87
		Funzionario	0	13	13
		Dirigente	0	4	4
	Totale		40	64	104

Uso degli elementi colonna per il controllo delle variabili nei file di dati

Nella finestra di dialogo Opzioni del Pannello di controllo SGO è possibile specificare gli elementi dimensione per le colonne, che verranno successivamente usati per creare le variabili nei file di dati generati. Questo processo è analogo ad effettuare il pivoting in Viewer.

Esempio: la procedura Frequenze produce una tabella statistica descrittiva in cui i dati statistici sono contenuti nelle righe; la procedure Descrittive invece produce una tabella statistica descritta in cui i dati statistici sono contenuti nelle colonne. Per poter

includere entrambe le tabelle in un file di dati in modo significativo, è necessario modificare la dimensione delle colonne di uno dei tipi di tabella.

Poiché entrambe le tabelle usano il nome “Statistiche” per la dimensione statistiche, è possibile inserire le statistiche della tabella della procedura Frequenze nella colonne selezionando semplicemente “Statistiche” (tra virgolette) nell’elenco dei nomi delle dimensioni nella finestra di dialogo Opzioni di SGO.

Figura 48-12
Finestra di dialogo Opzioni SGO

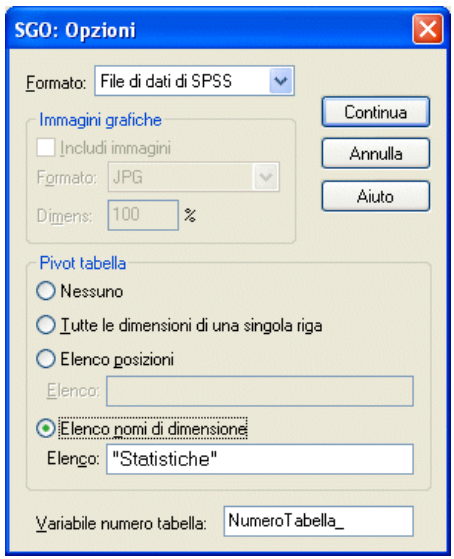


Figura 48-13
Inserimento di tipi di tabelle diversi in un file di dati tramite gli elementi della dimensione pivot

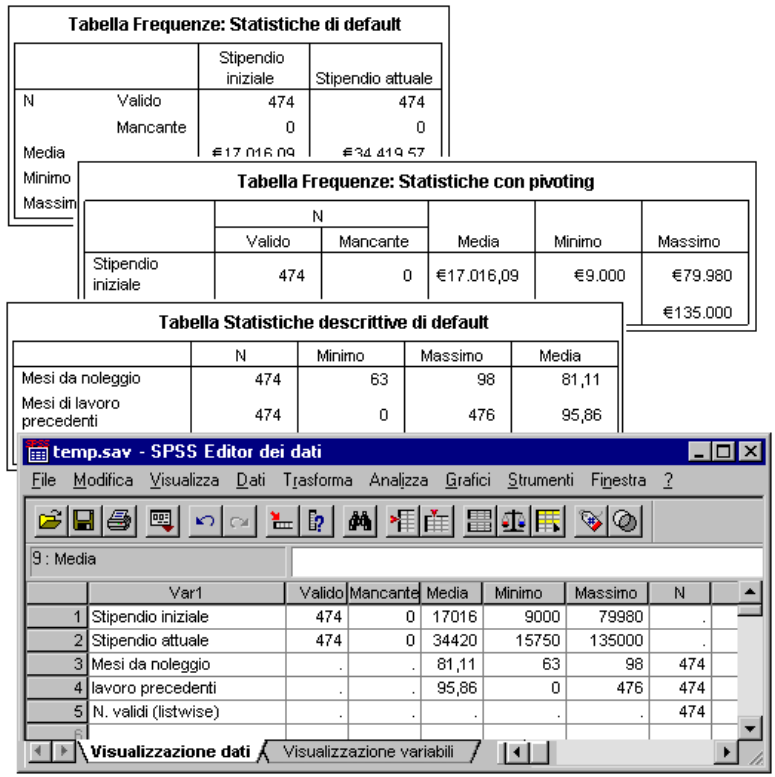


Tabella Frequenze: Statistiche di default

		Stipendio iniziale	Stipendio attuale
N	Valido	474	474
	Mancante	0	0
Media		€17.016,09	€34.419,57
Minimo			
Massim			

Tabella Frequenze: Statistiche con pivoting

	N		Media	Minimo	Massimo
	Valido	Mancante			
Stipendio iniziale	474	0	€17.016,09	€9.000	€79.980
					€135.000

Tabella Statistiche descrittive di default

	N	Minimo	Massimo	Media
Mesi da noleggio	474	63	98	81,11
Mesi di lavoro precedenti	474	0	476	95,86

temp.sav - SPSS Editor dei dati

File Modifica Visualizza Dati Trasforma Analizza Grafici Strumenti Finestra ?

9 : Media

	Var1	Valido	Mancante	Media	Minimo	Massimo	N
1	Stipendio iniziale	474	0	17016	9000	79980	.
2	Stipendio attuale	474	0	34420	15750	135000	.
3	Mesi da noleggio	.	.	81,11	63	98	474
4	lavoro precedenti	.	.	95,86	0	476	474
5	N. validi (listwise)	.	.	.	.	.	474

Visualizzazione dati | Visualizzazione variabili

Alcune delle variabili avranno dati mancanti poiché le strutture delle tabelle non sono 'esattamente le stesse dei dati statistici contenuti nelle colonne.

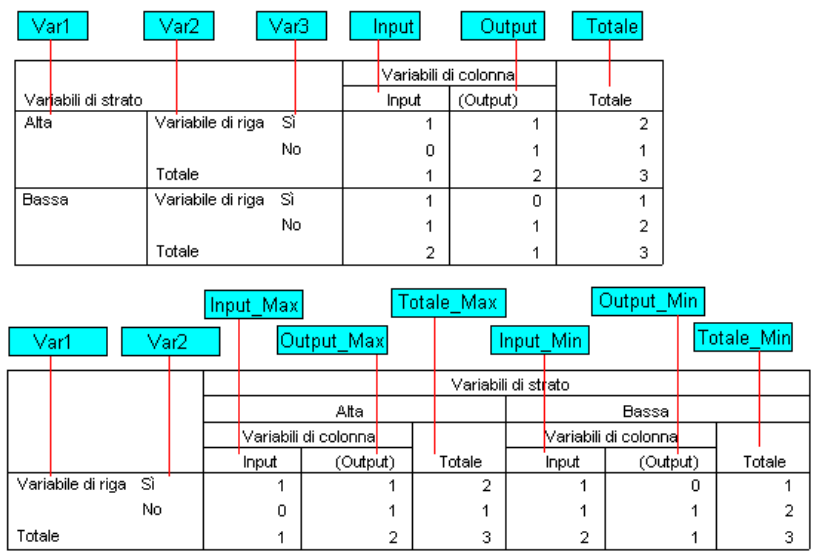
Nomi di variabili nei file dati generati da SGO

SGO genera nomi di variabili validi ed univoci sulla base delle etichette delle colonne:

- Agli elementi riga e strato vengono assegnati nomi di variabii generiche - ovvero il prefisso *Var* seguito da un numero progressivo.

- I caratteri non consentiti per 'i nomi delle variabili (come spazi e parentesi) vengono eliminati. Esempio: "Questa etichetta (colonna) " diventa una variabile denominata *QuestaEtichettaColonna*.
- Se l'etichetta inizia con un carattere ammesso per i nomi delle variabili ma non utilizzabile come primo carattere (ad esempio un numero), "@" viene inserito come prefisso. Esempio "2°" diventa una variabile denominata *@2°*.
- I trattini di sottolineatura o i punti alla fine delle etichette vengono cancellati dai nomi delle variabili risultanti. I trattini di sottolineature alla fine delle variabili generate automaticamente *Command_*, *Subtype_* e *Label_* non vengono rimossi).
- Se la dimensione colonna contiene più elementi, i nomi delle variabili vengono creati combinando le etichette delle categorie con i trattini di sottolineato tra le etichette delle categorie stesse. Le etichette dei gruppi non vengono incluse. Esempio. se *VarB* è nidificata sotto a *VarA* nelle colonne, le variabili risultanti sono *CatA1_CatB1*, non *VarA_CatA1_VarB_CatB1*.

Figura 48-14
Nomi di variabili creati dagli elementi della tabella



Struttura della tabella OXML

Output XML (OXML) è un XML conforme allo schema *spss-output*. Per una descrizione dettagliata dello schema, vedere *SPSSOutputXML_schema.htm* nella cartella *help\main* della cartella di installazione di SPSS.

- Il comando SGO e gli identificatori dei sottotipi vengono usati come valori del comando e attributi subType in OXML. Esempio:

```
<command text="Frequencies" command="Frequencies"...>
 <pivotTable text="Gender" label="Gender" subType="Frequencies"...>
```

- I valori degli attributi command e subType non sono influenzati dalla lingua dell'output o dalle impostazioni di visualizzazione impostate per i nomi/le etichette delle variabili o le etichette dei valori/del valore.
- XML distingue tra lettere minuscole e maiuscole. Il valore dell'attributo subType di "frequenze" *non* è uguale al valore dell'attributo subType di "Frequenze."
- Tutte le informazioni visualizzate nella tabella sono contenute nei valori degli attributi in OXML. A livello di cella singola, OXML è costituito da due elementi "vuoti" che contengono gli attributi ma nessun "contenuto" oltre a quello dei valori degli attributi.
- La struttura della tabella in OXML è rappresentata riga per riga; gli elementi che rappresentano le colonne sono nidificati nelle righe mentre le singole celle sono nidificate negli elementi delle colonne:

```
<pivotTable...>
 <dimension axis='row'...>
  <dimension axis='column'...>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
   <category...>
    <cell text='...' number='...' decimals='...'/>
   </category>
  </dimension>
 </dimension>
   ...
 </pivotTable>
```

L'esempio precedente è una rappresentazione semplificata della struttura che mostra le relazioni discendente/antenato di questi elementi, ma non è necessariamente una relazione padre/figlio a causa dei livelli di elementi nidificati.

La figura che segue mostra una tabella semplice della frequenza e la rappresentazione d'output completa della tabella in XML.

Figura 48-15
Tabella della frequenza semplice

Sesso

		Frequenza	Percentuale	Percentuale valida	Percentuale cumulata
Validi	Femmina	216	45,6	45,6	45,6
	Maschio	258	54,4	54,4	100,0
	Totale	474	100,0	100,0	

Figura 48-16
Output XML della tabella della frequenza semplice

```
<?xml version="1.0" encoding="UTF-8" ?>
<outputTreeoutputTree xmlns="http://xml.spss.com/spss/oms"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 xsi:schemaLocation="http://xml.spss.com/spss/oms
 http://xml.spss.com/spss/oms/spss-output-1.0.xsd">
 <command text="Frequencies" command="Frequencies"
 displayTableValues="label" displayOutlineValues="label"
 displayTableVariables="label" displayOutlineVariables="label">
 <pivotTable text="Gender" label="Gender" subType="Frequencies"
  varName="gender" variable="true">
  <dimension axis="row" text="Gender" label="Gender"
   varName="gender" variable="true">
   <group text="Valid">
    <group hide="true" text="Dummy">
     <category text="Female" label="Female" string="f"
      varName="gender">
      <dimension axis="column" text="Statistics">
       <category text="Frequency">
        <cell text="216" number="216"/>
       </category>
       <category text="Percent">
        <cell text="45.6" number="45.569620253165" decimals="1"/>
```

```
    </category>
    <category text="Valid Percent">
     <cell text="45.6" number="45.569620253165" decimals="1"/>
    </category>
    <category text="Cumulative Percent">
     <cell text="45.6" number="45.569620253165" decimals="1"/>
    </category>
   </dimension>
  </category>
  <category text="Male" label="Male" string="m" varName="gender">
   <dimension axis="column" text="Statistics">
    <category text="Frequency">
     <cell text="258" number="258"/>
    </category>
    <category text="Percent">
     <cell text="54.4" number="54.430379746835" decimals="1"/>
    </category>
    <category text="Valid Percent">
     <cell text="54.4" number="54.430379746835" decimals="1"/>
    </category>
    <category text="Cumulative Percent">
     <cell text="100.0" number="100" decimals="1"/>
    </category>
   </dimension>
  </category>
 </group>
 <category text="Total">
  <dimension axis="column" text="Statistics">
   <category text="Frequency">
    <cell text="474" number="474"/>
   </category>
   <category text="Percent">
    <cell text="100.0" number="100" decimals="1"/>
   </category>
   <category text="Valid Percent">
    <cell text="100.0" number="100" decimals="1"/>
   </category>
  </dimension>
 </category>
</group>
</dimension>
</pivotTable>
```

```
 </command>
</outputTree>
```

L'esempio mostra che anche una piccola tabella semplice può produrre una grande quantità di codice XML. Ciò' è dovuto in parte al fatto che il codice XML contiene informazioni non immediatamente visibili nella tabella originale, ossia informazioni che possono non essere disponibili nella tabella originale oltre ad una certa quantità di informazioni ridondanti.

- I contenuti della tabella che vengono visualizzati nella tabella pivot di Viewer sono contenuti negli attributi di testo. Esempio:

```
<command text="Frequenze" command="Frequenze"...>
```

- Gli attributi di testo sono influenzati sia dalla lingua dell'output che dalle impostazioni di visualizzazione dei nomi/delle etichette delle variabili e delle etichette dei valori/del valore. Nell'esempio, il valore dell'attributo testo varia a seconda della lingua di output,mentre il valore dell'attributo del comando rimane sempre lo stesso indipendentemente dalla lingua di output.
- Mentre le variabili o i valori vengono usati per le etichette di righe o colonne, in XML contengono un attributo testo ed uno o più valori di attributi aggiuntivi. Esempio:

```
<dimension axis="row" text="Gender" label="Gender" varName="gender">
  ...<category text="Female" label="Female" string="f" varName="gender">
```

- Nel caso di una variabile numerica, ci sarebbe un attributo numero anziché stringa. L'attributo etichetta è presente solo se per la variabile o i valori sono stati definite etichette.
- Gli elementi <cell> contengono i valori delle celle per i numeri che contengono l'attributotesto e uno o più valori attributo. Esempio:

```
<cell text="45.6" number="45.569620253165" decimals="1"/>
```

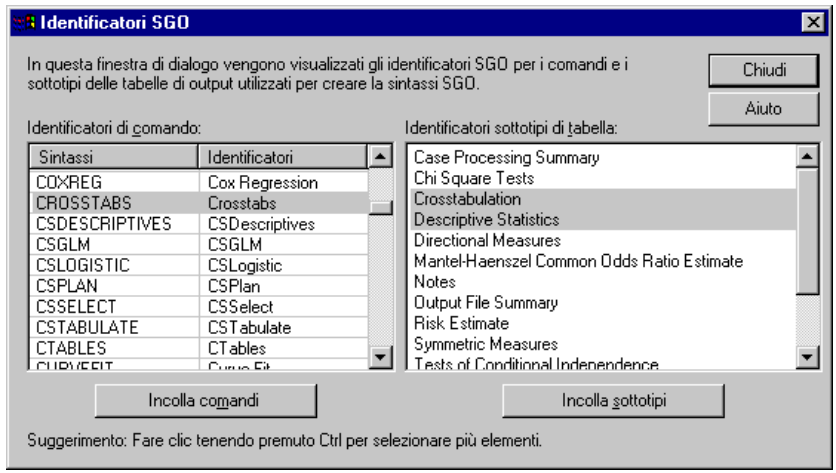
L'attributo numero rappresenta il valore numero effettivo non arrotondato, mentre l'attributo decimali indica il numero di posizioni decimali visualizzato nella tabella.

- Poiché le colonne sono nidificate nelle righe, l'elemento categoria che identifica ciascuna colonna viene ripetuto per tutte le righe. Esempio: poiché le statistiche sono visualizzate nelle colonne, l'elemento <category text="Frequenza"> appare tre volte in XML: una volta per la riga Maschio, una volta per la riga Femmina e una volta per la riga dei totali.

Identificatori SGO

La finestra di dialogo Identificatori SGO semplifica l'utilizzo della sintassi dei comandi di SGO poiché consente di incollare il comando selezionato e gli identificatori del sottotipo in una finestra della sintassi dei comandi.

Figura 48-17
Finestra di dialogo Identificatori SGO



Per utilizzare la finestra di dialogo Identificatori SGO

▶ Dai menu, scegliere:

Strumenti
  Identificatori SGO...

▶ Selezionare uno o più degli identificatori di comandi o di sottotipi. Per selezionare più identificatori in ogni elenco, fare clic su di essi tenendo premuto Ctrl.

- Fare clic su Incolla comandi e/o su Incolla sottotipi.
  - L'elenco dei sottotipi disponibili varia in base ai comandi correntemente selezionati. Se sono selezionati più comandi, l'elenco dei sottotipi è rappresentato dall'unione di tutti i sottotipi disponibili per ognuno dei comandi selezionati. Se non sono selezionati comandi, l'elenco includerà tutti i sottotipi.
  - Gli identificatori vengono incollati nella finestra della sintassi dei comandi designata in corrispondenza della posizione del cursore corrente. Se non sono aperte finestre della sintassi dei comandi, ne verrà aperta automaticamente una nuova.
  - Ogni identificatore di comando e/o sottotipo viene racchiuso tra virgolette quando viene incollato, poiché la sintassi del comando `SGO` le richiede.
  - Gli elenchi di identificatori per le parole chiave `COMMANDS` e `SUBTYPES` devono essere racchiusi tra parentesi quadre, come è illustrato nell'esempio seguente:

```
/IF COMMANDS=['Crosstabs' 'Descriptives']
    SUBTYPES=['Crosstabulation' 'Descriptive Statistics']
```

Copia degli identificatori SGO dalla vista riassuntiva del Viewer

possibile copiare e incollare gli identificatori di comandi e sottotipi dalla vista riassuntiva del Viewer.

- Nella vista riassuntiva fare clic con il pulsante destro del mouse su una voce.
- Selezionare Copia identificatore comandi SGO oppure Copia sottotipo tabelle SGO.

  Questo metodo si differenzia dall'utilizzo della finestra Identificatori SGP per il fatto che l'identificatore copiato non viene incollato automaticamente in una finestra della sintassi dei comandi, ma viene semplicemente copiato negli Appunti, dai quali è possibile incollarlo in qualsiasi posizione. Dato che i valori degli identificatori di comandi e sottotipi sono identici ai comandi e ai valori degli attributi di sottotipo corrispondenti nell'output in formato XML (OXML), questo metodo di copia e incolla può risultare utile quando si scrivono trasformazioni XSLT.

Copia di etichette SGO

Anzichè gli identificatori, è possibile copiare le etichette e utilizzarle con la parola chiave `LABELS`. Le etichette consentono di differenziare più grafici o più tabelle dello stesso tipo nei quali il testo dell'output rappresenta un attributo dell'oggetto di output specifico, ad esempio i nomi o le etichette delle variabili. Il testo delle etichette può tuttavia essere influenzato da numerosi fattori:

- Se è attiva la distinzione per gruppi, è possibile che all'etichetta venga aggiunta l'identificazione del gruppo di file distinto in base alla variabile categoriale.
- Le etichette che includono informazioni sulle variabili o i valori sono influenzate dalle impostazioni di visualizzazione per i nomi e le etichette delle variabili e per i valori e le etichette di valori nella vista riassuntiva (menu Modifica, Opzioni, scheda Etichette di output).
- Le etichette sono influenzate dall'impostazione corrente della lingua di output (menu Modifica, Opzioni, scheda Generale).

Per copiare le etichette SGO

▶ Nella vista riassuntiva fare clic con il pulsante destro del mouse su una voce.

▶ Selezionare Copia etichetta SGO.

Allo stesso modo degli identificatori di comandi e di sottotipi, le etichette devono essere racchiuse tra virgolette e l'elenco completo deve essere racchiuso tra parentesi quadre, come è illustrato nell'esempio seguente:

```
/IF LABELS=['Employment Category' 'Education Level']
```

Appendice

A

Database Access Administrator

Database Access Administrator è un'utilità studiata per semplificare l'utilizzo di sorgenti dati complesse o di grandi dimensioni con la Creazione guidata di query. Database Access Administrator consente agli utenti e agli amministratori di personalizzare la sorgente dati nei modi seguenti:

- Creare alias per le tabelle e i campi del database.
- Creare nomi di variabili per i campi.
- Nascondere tabelle e campi non validi.

Database Access Administrator non modifica il database, ma genera file che contengono tutte le informazioni. Questi file fungono da *visualizzazioni* database.

Database Access Administrator può essere utilizzato per specificare fino a tre visualizzazioni per database: livello Società, livello Reparto e livello Privato. Database Access Administrator e Creazione guidata di query riconoscono tali file in base ai seguenti nomi:

- Livello Società: *dba01.inf*
- Livello Reparto: *dba02.inf*
- Livello Privato: *dba03.inf*

Ciascun file contiene informazioni specifiche del livello su qualunque numero di sorgenti dati. Ad esempio, il file *dba03.inf* può contenere informazioni personali per un database degli account societari, il database degli orari e un database utilizzato per gestire la propria raccolta di CD.

All'apertura, Database Access Administrator effettua una ricerca dei file nel percorso di sistema e visualizza automaticamente le informazioni relative alle tre visualizzazioni delle sorgenti dati configurate.

Origini e priorità. Quando viene utilizzato, Creazione guidata di query presenta la visualizzazione di livello inferiore della sorgente dati trovata nel percorso del sistema. La gerarchia dei livelli, dal più alto al più basso è: Società, Reparto e Privato. Il file di ciascun livello contiene informazioni relative a tutte le sorgenti dati di quel livello. Ad esempio, il reparto marketing disporrà di un file, *dba02.inf*, contenente le informazioni di alias per tutte le visualizzazioni del database definite per il reparto marketing. Ciascuna persona del reparto marketing disporrà di un proprio file *dba03.inf* contenente visualizzazioni personalizzate di tutti i database utilizzati.

In Database Access Administrator, gli alias, i nomi delle variabili e le disposizioni di nascondere elementi vengono trasmessi dall'alto verso il basso.

Esempi:

- Se la tabella Aree è nascosta a livello Società, diventa invisibile anche ai livelli Reparto e Privato. La tabella non verrà visualizzata nella Creazione guidata di query.
- Il campo *CATLAV* della tabella Vendite Dipendente non ottiene un alias a livello Società, mentre lo riceve come Categorie lavorative a livello Reparto. Appare come Categorie lavorative anche a livello Privato. Inoltre, se la tabella Vendite Dipendente riceve un alias come Informazioni dipendente a livello Privato, il campo originale (*VenditeDipendente.CATLAV*) verrà visualizzato nella Creazione guidata di query come *'Informazioni dipendente'.'Categorie lavorative'*.

Per avviare Database Access Administrator, eseguire il file *spssdbca.exe*, installato nella directory SPSS. Per ulteriori informazioni su Database Access Administrator, vedere il relativo Aiuto in linea.

Appendice

B

Personalizzazione dei documenti HTML

È possibile aggiungere automaticamente codice HTML personalizzato ai documenti esportati in formato HTML, inclusi:

- Titoli di documenti HTML
- Specifiche del tipo di documento
- Metatag e codice dello script (ad esempio, JavaScript)
- Testo visualizzato prima e dopo l'output esportato

Per aggiungere codice HTML personalizzato a documenti di output esportati

▶ Aprire il file *htmlfram.txt* (che si trova nella directory in cui è installato SPSS) con un editor di testi.

▶ Sostituire i commenti nei "campi" sulle righe posti fra le doppie parentesi angolari aperte (<<) con il testo o il codice HTML da inserire nei documenti HTML esportati.

▶ Salvare il file come file di testo.

Nota: se si modifica il nome o la posizione del file di testo, è necessario modificare il registro di sistema per poter utilizzare il file per personalizzare l'output HTML esportato.

Contenuto e formato del file di testo per HTML personalizzato

Il codice HTML da aggiungere automaticamente ai documenti HTML deve essere specificato in un file di testo semplice contenente sei campi, ciascuno delimitato da due parentesi angolari aperte (<<) poste sulla riga precedente:

<<
Il testo o il codice da inserire nella parte superiore del documento prima della specifica `<HTML>`. Ad esempio, è possibile inserire commenti che includono specifiche sul tipo di documento.

<<
Testo utilizzato come titolo del documento (visualizzato nella barra del titolo)

<<
Metatag o codice dello script (ad esempio, JavaScript)

<<
Codice HTML che modifica l'identificatore `<BODY`, ad esempio un codice che specifica il colore di sfondo

<<
Testo e/o codice HTML inserito dopo l'output esportato, ad esempio informazioni sul copyright

<<
Testo e/o codice HTML inserito prima dell'output esportato, ad esempio nome della società, logo e così via.

Per utilizzare un file o una posizione diversa per il codice HTML personalizzato

Se si modifica il nome o la posizione del file *htmlfram.txt* è necessario modificare il registro di sistema per utilizzare il file nell'output HTML personalizzato.

- Nel menu Start di Windows, scegliere Esegui, digitare regedit e fare clic su OK.

- Nel riquadro sinistro dell'Editor del registro di configurazione, selezionare:

```
HHKEY_CURRENT_USER
  Software
    SPSS
      SPSS per Windows
        14.0
          Spsswin
```

- Nel riquadro destro fare doppio clic sulla stringa HTMLFormatFile.
- In Dati valore specificare il percorso completo e il nome del file di testo contenente le specifiche del codice HTML personalizzato (ad esempio, *c:\myfiles\htmlstuf.txt*).

File di testo di esempio per HTML personalizzato

```
<<
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2//EN"<
< NVI, Inc.
<<
<META NAME="keywords" CONTENT="gizmos, gadgets, gimcracks">
<<
bgcolor="#FFFFFF"
<<
<H4 align=center>This page made possible by...
<br><br>
<IMG SRC="spss2.gif" align=center</H4>
<<
<h2 align=center>NVI Sales</h2>
<h3 align=center>Regional Data</h3>
```

Sorgente HTML di esempio per HTML personalizzato

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2//EN">
<HTML>
<HEAD>
<TITLE>
NVI Sales, Inc.
</TITLE>
<META NAME="keywords" CONTENT="gizmos, gadgets, gimcracks">
```

```
</HEAD>
<BODY bgcolor="#FFFFFF">
<h2 align=center>NVI Sales</h2>
<h3 align=center>Regional Data</h3>

[Esportazione dell'output]

<H4 align=center>This page made possible by...
<br><br>
<IMG SRC="spss2.gif" align=center></H4>
</BODY>
</HTML>
```

Indice

Access (Microsoft), 33
affidabilità di divisione a metà
  in analisi di affidabilità, 587, 589
affidabilità di Spearman-Brown
  in analisi di affidabilità, 589
aggiunta di etichette di gruppo, 304
aggregazione dei dati, 221
  funzioni di aggregazione, 224
  nomi di variabili ed etichette, 225
aiuto contestuale, 306
  ricerca delle definizioni di etichetta nelle tabelle pivot, 306
aiuto in linea, 17
  Analisi guidate, 15
alfa di Cronbach
  in analisi di affidabilità, 587, 589
algoritmi, 17
allineamento, 106, 261, 327, 632
  delle celle, 327
  nell'Editor dei dati, 106
  output, 261, 632
allocazione della memoria
  in analisi Cluster TwoStep, 511
analisi a risposta multipla
  Risposte multiple: Frequenze, 564
  Risposte multiple: Tavole di contingenza, 566
  tabelle di frequenza, 564
  tavole di contingenza, 566
analisi cluster
  Cluster con metodo delle K-medie, 525
  Cluster gerarchica, 517
  efficienza, 529
Analisi Cluster TwoStep, 507
  di scala, 515
  grafici, 514
  opzioni, 511
  salvataggio in un file di lavoro, 515
  salvataggio in un file esterno, 515
analisi dei dati, 14
  passaggi fondamentali, 14
analisi della varianza
  in ANOVA univariata, 411
  in medie, 388
  in regressione lineare, 468
  in stima di curve, 473
analisi delle componenti principali, 493, 499
Analisi di affidabilità, 587
  coefficiente di correlazione intraclasse, 589
  correlazione e covarianze inter-item, 589
  descrittive, 589
  di scala, 587, 589
  esempio, 587
  Kuder-Richardson 20, 589
  opzioni aggiuntive del comando, 591
  tabella ANOVA, 589
  test di additività di Tukey, 589
  *T*quadrato di Hotelling , 589
Analisi discriminante, 481
  analisi di casi esclusi, 676
  coefficienti di funzione, 485
  criteri, 486
  definizione di intervalli, 484
  di scala, 481, 485
  Distanza di Mahalanobis, 486

esempio, 481
esportazione di informazioni dei modelli, 490
grafici, 488
Lambda di Wilks, 486
matrice di covarianza, 488
matrici, 485
metodi discriminanti, 486
metodi stepwise, 481
opzioni aggiuntive del comando, 490
opzioni di visualizzazione, 486, 488
probabilità a priori, 488
salvataggio di variabili di classificazione, 490
selezione di casi, 484
statistiche descrittive, 485
V di Rao , 486
valori mancanti, 488
variabili di raggruppamento, 481
variabili indipendenti, 481
Analisi fattoriale, 493
analisi di casi esclusi, 676
cenni generali, 493
convergenza, 499, 501
descrittive, 498
di scala, 493, 498
esempio, 493
formato di visualizzazione dei coefficienti, 503
grafici dei pesi fattoriali, 501
metodi di estrazione, 499
metodi di rotazione, 501
opzioni aggiuntive del comando, 503
punteggi fattoriali, 502
selezione di casi, 497
valori mancanti, 503
Analisi guidate, 15
analisi serie storiche
previsione, 478
previsione di casi, 478
anni, 645
valori a due cifre, 645
ANOVA
in ANOVA univariata, 411
in GLM univariato, 421
in medie, 388
modello, 425
ANOVA univariata, 411
confronti multipli, 415
contrasti, 414
contrasti polinomiali, 414
di scala, 418
opzioni, 418
opzioni aggiuntive del comando, 419
test Post Hoc, 415
valori mancanti, 418
variabili fattore, 411
anticipo, 202
apertura di file, 27–32, 51
file di dati, 27–28
file di dati testuali, 51
file di dBASE, 27, 30
file di Excel, 27, 29
file di fogli elettronici, 27, 30
File di Lotus 1-2-3, 27
File Stata, 31
file SYSTAT, 27
file tab delimitati, 27
asimmetria
in cubi OLAP, 394
in descrittive, 355
in esplora, 363
in frequenze, 348
in medie, 388
in Report: Riepiloghi per colonne, 582

in Report: Riepiloghi per righe, 575
in Riassumi, 382
associazione lineare
in tavole di contingenza, 372
attributi di variabili, 106, 108
copiare e incollare, 106, 108
automazione OLE, 675, 689, 691–692, 694–695
cenni generali, 689, 691–692, 695
convenzioni di denominazione delle variabili, 691
metodi, 694
proprietà, 694
uso degli script, 675
utilizzo di oggetti, 689, 692, 695
autovalori
in analisi fattoriale, 498–499
in regressione lineare, 468

barra di stato, 10
mostrare, 10
nascondere, 10
barre degli strumenti, 652, 654–656
creazione, 652, 655
creazione di nuovi strumenti, 656
modifica delle icone, 656
mostrare e nascondere, 652
personalizzazione, 652, 655
visualizzazione in finestre diverse, 654
Bonferroni
in ANOVA univariata, 415
in GLM, 431
bordi, 293, 320, 322
Bozza, 293
visualizzazione dei bordi nascosti, 322
bordi delle celle, 293
Bozza, 293
Bozza, 291–293, 298–300, 630, 633
bordi delle celle, 293
bordi delle colonne, 293
bordi delle righe, 293
controllo della visualizzazione predefinita dell'output, 630
formato di output, 293
impostazione del Viewer di default, 292
linee continue, 293
modifica di caratteri, 298
opzioni di visualizzazione, 633
salvataggio dell'output, 300
stampa, 298–299

C di Dunnett
in ANOVA univariata, 415
in GLM, 431
calcolo di variabili, 161
calcolo di nuove variabili stringa, 164
campionamento
campione casuale, 230
campione casuale, 41
file di database, 41
selezione, 230
seme dei numeri casuali, 165
campioni dipendenti, 551, 558
caratteri, 115, 264, 298, 323
colori, 323
delle celle, 323
in bozza, 298
nell'Editor dei dati, 115
nella vista riassuntiva, 264
casi, 112, 114, 234
inserimento di nuovi casi, 112
ordinamento, 212
ponderazione, 232

ricerca di duplicati, 146
ricerca nell'Editor dei dati, 114
ristrutturazione in variabili, 234
selezione di sottoinsiemi, 227, 229, 231
casi duplicati (record)
ricerca e funzioni di filtro, 146
casi esclusi, 676
analisi, 676
casi filtrati, 115
nell'Editor dei dati, 115
categoria di riferimento
in GLM, 428
celle nelle tabelle pivot, 313, 319, 323–324, 326–329, 331
allineamento, 327
caratteri, 323
contorni, 329
formati, 319
formati di valori, 326
larghezze, 324
margini, 328
modifica di testo, 331
mostrare, 311
nascondere, 311
ombreggiatura, 329
cenni generali, 745
centratura dell'output, 261, 632
chi-quadrato, 534
associazione lineare, 372
correzione di continuità di Yates, 372
di scala, 537
in tavole di contingenza, 372
intervallo atteso., 536
opzioni, 537
Pearson, 372
per l'indipendenza, 372
rapporto di verosimiglianza, 372
test esatto di Fisher, 372
test per un campione, 534
valori attesi, 536
valori mancanti, 537
chi-quadrato del rapporto di verosimiglianza
in tavole di contingenza, 372
Chi-quadrato di Pearson
in tavole di contingenza, 372
classifica di casi, 175
percentili, 177
punteggi esponenziali, 177
ranghi frazionari, 177
valori a pari merito, 178
classificazione
nella Curva ROC, 619
Cluster con metodo delle K-medie, 525
cenni generali, 525
cluster di appartenenza, 531
criteri di convergenza, 530
di scala, 525, 531
distanze tra cluster, 531
efficienza, 529
esempi, 525
iterazioni, 530
metodi, 525
opzioni aggiuntive del comando, 532
salvataggio di informazioni del cluster, 531
valori mancanti, 531
Cluster gerarchica, 517
cluster di appartenenza, 522–523
dendrogrammi, 523
di scala, 517, 522
esempio, 517
grafici a stalattite, 523
matrici delle distanze, 522

metodi di raggruppamento, 521
misure di distanza, 521
misure di similarità, 521
opzioni aggiuntive del comando, 524
orientamento del grafico, 523
programmi di agglomerazione, 522
raggruppamento dei casi in cluster, 517
salvataggio di nuove variabili, 523
trasformazione di misure, 521
trasformazione di valori, 521
variabili, 517
coefficiente alfa
in analisi di affidabilità, 587, 589
coefficiente di contingenza
in tavole di contingenza, 372
coefficiente di correlazione dei ranghi
in correlazioni bivariate, 439
coefficiente di correlazione di Spearman
in correlazioni bivariate, 439
in tavole di contingenza, 372
coefficiente di correlazione intraclasse
in analisi di affidabilità, 589
coefficiente di correlazione *r*
in correlazioni bivariate, 439
in tavole di contingenza, 372
coefficiente di dispersione (COD)
in Statistiche di rapporto, 603
coefficiente di incertezza
in tavole di contingenza, 372
coefficiente di rischio
in tavole di contingenza, 372
coefficiente di variazione (COV)
in Statistiche di rapporto, 603
coefficienti beta
in regressione lineare, 468
coefficienti di regressione.
in regressione lineare, 468
colonna totale
nei report, 583
colonne, 324, 331
modifica della larghezza nelle tabelle pivot, 324
selezione nelle tabelle pivot, 331
colori nelle tabelle pivot, 320, 323, 329
bordi, 320
carattere, 323
primo piano di una cella, 329
sfondo di una cella, 329
comando EXECUTE
copiato dalle finestre di dialogo, 342
compressione delle categorie, 149
condizioni, 163
confronti multipli
in ANOVA univariata, 415
confronti multipli post hoc, 415
confronto di gruppi
in cubi OLAP, 397
confronto di variabili
in cubi OLAP, 397
conteggio atteso
in tavole di contingenza, 376
conteggio delle occorrenze, 167
conteggio osservato
in tavole di contingenza, 376
contrasti
in ANOVA univariata, 414
in GLM, 428
contrasti di deviazione
in GLM, 428
contrasti di Helmert
in GLM, 428

contrasti differenza
in GLM, 428
contrasti polinomiali
in ANOVA univariata, 414
in GLM, 428
contrasti ripetuti
in GLM, 428
contrasti semplici
in GLM, 428
controllo pagina
nei report di riepilogo per colonne, 584
nei report di riepilogo per righe, 577
convergenza
in analisi fattoriale, 499, 501
nel cluster con metodo delle K-medie, 530
copia, 261, 266
grafici, 266
output, 261
tabelle pivot, 266
coppie di variabili, 234
creazione, 234
correlazione di Pearson
in correlazioni bivariate, 439
in tavole di contingenza, 372
correlazioni
di ordine zero, 448
in correlazioni bivariate, 439
in Correlazioni parziali, 445
in tavole di contingenza, 372
Correlazioni bivariate
coefficienti di correlazione, 439
di scala, 442
livello di significatività, 439
opzioni, 442
opzioni aggiuntive del comando, 443
valori mancanti, 442
correlazioni di ordine zero
in Correlazioni parziali, 448
Correlazioni parziali, 445
correlazioni di ordine zero, 448
di scala, 448
in regressione lineare, 468
opzioni, 448
opzioni aggiuntive del comando, 448
valori mancanti, 448
correzione di continuità di Yates
in tavole di contingenza, 372
costruzione di termini, 426
Cubi OLAP, 391
di scala, 394
titoli, 398
curtosi
in cubi OLAP, 394
in descrittive, 355
in esplora, 363
in frequenze, 348
in medie, 388
in Report: Riepiloghi per colonne, 582
in Report: Riepiloghi per righe, 575
in Riassumi, 382
Curva ROC, 619
statistiche e grafici, 622

d
in tavole di contingenza, 372
d di Somers
in tavole di contingenza, 372
Data List, 78
confronto con il comando Get Data, 78
dati categoriali, 129
conversione di dati di intervallo in categorie discrete, 149

dati delimitati da spazi, 51
dati delimitati da virgole, 51
dati di serie storiche
  creazione di nuove variabili di serie storica, 201
  definizione di variabili di data, 199
  funzioni di trasformazione, 202
  sostituzione di valori mancanti, 203
  trasformazioni di dati, 198
Dati Dimensions, 62
dati in formato CSV, 51
dati pesati, 255
  e file di dati ristrutturati, 255
debug degli script, 704–705
  analisi degli script, 704
  punti di interruzione, 704
  riquadro di debug, 705
definizione di variabili, 98, 100, 103, 105–106, 108, 124
  applicazione di un dizionario dati, 136
  copiare e incollare attributi, 106, 108
  etichette dei valori, 103, 124
  etichette di variabili, 103
  modelli, 106, 108
  tipi di dati, 100
  valori mancanti, 105
dendrogrammi
  in cluster gerarchica, 523
Descrittive, 353
  di scala, 355
  opzioni aggiuntive del comando, 356
  ordine di visualizzazione, 355
  salvataggio di punteggi *z*, 353
deviazione media assoluta (AAD)
  in Statistiche di rapporto, 603
deviazione standard
  in cubi OLAP, 394
  in descrittive, 355
  in esplora, 363
  in frequenze, 348
  in GLM univariato, 436
  in medie, 388
  in Report: Riepiloghi per colonne, 582
  in Report: Riepiloghi per righe, 575
  in Riassumi, 382
  in Statistiche di rapporto, 603
diagnostica di collinearità
  in regressione lineare, 468
diagnostiche per casi
  in regressione lineare, 468
dichiarazioni di variabili, 690–691
  convenzioni di denominazione, 691
  negli script, 690–691
didascalie, 332
  aggiunta a una tabella, 332
differenza, 202
differenza in beta
  in regressione lineare, 465
differenza meno significativa
  in ANOVA univariata, 415
  in GLM, 431
differenza meno significativa (LSD) di Fisher
  in GLM, 431
differenza nel valore adattato
  in regressione lineare, 465
differenza significativa di Tukey
  in ANOVA univariata, 415
  in GLM, 431
differenza stagionale, 202
differenze tra gruppi
  in cubi OLAP, 397
differenze tra variabili
  in cubi OLAP, 397

differenziale di prezzo (PRD)
in Statistiche di rapporto, 603
dimensionamento di grafici esportati, 277
dimensioni, 264
nella vista riassuntiva, 264
directory temporanea, 630
impostazione della posizione in modalità locale, 630
variabile di ambiente SPSSTMPDIR, 630
distanza chi-quadrato
in distanze, 453
distanza City-Block
in distanze, 453
distanza di Chebychev
in distanze, 453
distanza di Cook
in GLM, 433
in regressione lineare, 465
distanza di Mahalanobis
in regressione lineare, 465
Distanza di Mahalanobis
in analisi discriminante, 486
distanza euclidea
in distanze, 453
distanza euclidea quadratica
in distanze, 453
distanza Minkowski
in distanze, 453
Distanze, 451
calcolo delle distanze tra casi, 451
calcolo delle distanze tra variabili, 451
di scala, 451
esempio, 451
in cluster gerarchica, 517
misure di dissimilarità, 453
misure di similarità, 454
opzioni aggiuntive del comando, 455
trasformazione di misure, 453–454
trasformazione di valori, 453–454
distinzione per gruppi, 226
divisione
divisione tra colonne del report, 583
divisore di finestre
Editor dei dati, 115
divisore di riquadri
Editor dei dati, 115
dizionario, 68–69
dizionario dati
applicazione da un altro file, 136
Dunnett (T3)
in ANOVA univariata, 415
in GLM, 431

Editor dei dati, 95, 98, 106, 108–115, 117, 651
allineamento, 106
casi filtrati, 115
definizione di variabili, 98
file di dati multipli aperti, 119
inserimento di dati, 108
inserimento di dati non numerici, 110
inserimento di dati numerici, 109
inserimento di nuove variabili, 113
inserimento di nuovi casi, 112
invio di dati ad altre applicazioni, 651
larghezza di colonna, 106
limitazioni relative ai valori, 110
modifica del tipo di dati, 114
modifica di dati, 111–112
opzioni di visualizzazione, 115
ricerca di casi, 114
spostamento delle variabili, 114
stampa, 117

viste/riquadri multipli, 115
Visualizzazione dati, 96
Visualizzazione variabili, 97
elenchi di destinazione, 627
elenchi di variabili, 627
riordinamento degli elenchi di destinazione, 627
elenco dei casi, 379
eliminazione all'indietro
in regressione lineare, 461
eliminazione dell'output, 260
eliminazione di comandi EXECUTE multipli nei file di sintassi, 342
errore standard
in descrittive, 355
in esplora, 363
in frequenze, 348
in GLM, 433, 436
nella Curva ROC, 622
errore standard dell'asimmetria
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
errore standard della curtosi
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
errore standard della media
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
esclusione dell'output da Viewer con SGO, 726
Esplora, 359
di scala, 363
grafici, 364
opzioni, 366
opzioni aggiuntive del comando, 366
trasformazioni di potenza, 365
valori mancanti, 366
esportazione dell'output, 270, 274–275, 662, 674
formato Excel, 270, 274
Formato PowerPoint, 270
formato Word, 270, 274
HTML, 274
pubblicazione sul Web, 674
SGO, 711
esportazione di dati, 651
aggiunta di comandi di menu per l'esportazione di dati, 651
esportazione di grafici, 270, 277–281, 659, 662
dimensioni del grafico, 277
produzione automatica, 659
eta
in medie, 388
in tavole di contingenza, 372
eta-quadrato
in GLM univariato, 436
in medie, 388
etichette, 304
eliminazione, 304
inserimento di etichette di gruppo, 304
vs. nomi di sottotipi in SGO, 718
etichette dei valori, 103, 110, 115, 124, 635
applicazione a più variabili, 130
copia, 130
in file di dati uniti, 218
inserimento di interruzioni di riga, 104
nell'Editor dei dati, 115
nella vista riassuntiva, 635
nelle tabelle pivot, 635
utilizzo per l'inserimento dei dati, 110
etichette di gruppo, 304

etichette di variabili, 103, 630, 635
  in file di dati uniti, 218
  inserimento di interruzioni di riga, 104
  nella vista riassuntiva, 635
  nelle finestre di dialogo, 11, 630
  nelle tabelle pivot, 635
é*V* di Cram
  in tavole di contingenza, 372
eventi di attivazione, 686
  script automatici, 686

F multiplo di Ryan-Einot-Gabriel-Welsch
  in ANOVA univariata, 415
  in GLM, 431
fattore di inflazione della varianza
  in regressione lineare, 468
fattorizzazione alfa, 499
fattorizzazione dell'asse principale, 499
fattorizzazione immagine, 499
file, 265
  aggiunta di un file di testo al Viewer, 265
  apertura, 27
file attivo, 78, 80
  creazione di un file attivo temporaneo, 80
  file attivo virtuale, 78
  inserimento nella cache, 80
file attivo temporaneo, 80
file BMP, 270, 277–278, 662
  esportazione di grafici, 270, 277–278, 662
file di database, 31–33, 36–37, 39–41, 44, 47, 49
  accesso a un database, 36
  campionamento casuale, 41
  conversione di variabili stringa in variabili numeriche, 47
  creazione di relazioni, 39
  definizione di criteri, 41
  definizione di variabili, 47
  espressioni logiche, 41
  join di tabelle, 39–40
  lettura, 31–33, 37
  Microsoft Access, 33
  proposizione Where, 41
  proprietà delle relazioni, 40
  protezione di database, 36
  query di parametri, 41, 44
  richiesta di valore, 44
  salvataggio di query, 49
  selezione di campi dei dati, 37
  selezione di una sorgente dati, 33
  sintassi SQL, 49
  verifica dei risultati, 49
file di dati, 27–28, 51, 68–69, 73–74, 76, 80, 88–90, 234
  aggiunta di commenti, 624
  apertura, 27–28
  capovolgere, 213
  Dimensions, 62
  file di dati multipli aperti, 119
  informazioni del dizionario, 68–69
  informazioni sul file, 68–69
  miglioramento delle prestazioni per file di grandi dimensioni, 80
  mrInterview, 62
  protezione, 77
  Quancept, 62
  Quanvert, 62
  ridenominazione, 122
  ristrutturazione, 234
  salvare, 69, 73–74
  salvataggio dell'output come file di dati in formato SPSS, 711
  salvataggio di sottoinsiemi di variabili, 76

server remoti, 88–90
testo, 51
trasposizione, 213
file di dati multipli aperti, 119
file di dBASE, 27, 30, 73–74
apertura, 27, 30
salvare, 73–74
file di Excel, 27, 29, 73–74, 651
aggiunta di un comando di menu per inviare dati a Excel, 651
apertura, 27, 29
salvare, 73–74
file di fogli elettronici, 27, 29–30, 77
apertura, 30
lettura degli intervalli, 29
lettura dei nomi di variabili, 29
scrittura di nomi di variabili, 77
File di Lotus 1-2-3, 27, 73–74, 651
aggiunta di un comando di menu per inviare dati a Lotus, 651
apertura, 27
salvare, 73–74
file EPS, 270, 277, 280, 662
esportazione di grafici, 270, 277, 280, 662
file giornale, 630
file in formato portabile
nomi di variabili, 74
file JPEG, 270, 277–278, 662
esportazione di grafici, 270, 277–278, 662
file modello
caricamento di modelli salvati nei punteggi dei dati, 206
file PICT, 270, 277–278, 662
esportazione di grafici, 270, 277–278, 662
file PNG, 270, 279
esportazione di grafici, 270, 279
file PostScript (encapsulated), 270, 280, 662
esportazione di grafici, 270, 280, 662
file SAS
apertura, 27
salvare, 73
File Stata, 31
apertura, 31
lettura, 27
salvare, 73
file SYSTAT, 27
apertura, 27
file tab delimitati, 27, 29, 51, 73–74, 77
apertura, 27
lettura dei nomi di variabili, 29
salvare, 73–74
scrittura di nomi di variabili, 77
file TIFF, 279
esportazione di grafici, 270, 277, 279, 662
file WMF, 270, 277, 281, 662
esportazione di grafici, 270, 277, 281, 662
finestra attiva, 8
finestra dello script, 680, 682, 695
menu Debug, 704
proprietà, 682
Scheda Immediate, 705
Scheda Loaded, 705
Scheda Stack, 705
Scheda Watch, 705
Visualizzatore oggetti, 695
finestra designata, 8
finestre, 6
finestra attiva, 8
finestra designata, 8
Finestre dell'Aiuto, 17
finestre di dialogo, 13, 625–626, 630, 699, 701
controlli, 12

definizione di insiemi di variabili, 625
icone delle variabili, 13
informazioni sulle variabili, 13
ordine di visualizzazione delle variabili, 630
riordinamento degli elenchi di destinazione, 627
selezione di variabili, 13
sottofinestre di dialogo, 12
sovrapposte, 10
specifiche facoltative, 12
utilizzo di insiemi di variabili, 626
utilizzo di script, 699, 701
visualizzazione di etichette di variabili, 11, 630
visualizzazione di nomi di variabili, 11, 630
formati di data, 100, 102, 645
anni a due cifre, 645
formati di input, 102
formati di valuta, 646
formati di valuta personalizzati, 100, 646
formati di visualizzazione, 102
formato colonne, 326
modifica nelle tabelle pivot, 326
formato dollaro, 100, 102
formato Excel
esportazione dell'output, 270, 274
formato file dati SPSS
trasmissione dell'output ad un file dati, 720, 726
formato file sav
trasmissione dell'output ad un file dati SPSS, 720, 726
formato fisso, 51
formato libero, 51
formato numerico, 100, 102
Formato PowerPoint
esportazione dell'output, 270
formato punto, 100, 102
formato stringa, 100
formato virgola, 100, 102
formato Word
esportazione dell'output, 270, 274
formattazione, 293
colonne nei report, 574
output bozza, 293
Frequenze, 345
di scala, 348
formati, 351
grafici, 350
ordine di visualizzazione, 351
soppressione di tabelle, 351
frequenze dei cluster
in analisi Cluster TwoStep, 515
funzioni, 164, 696
trattamento dei valori mancanti, 165

gamma
in tavole di contingenza, 372
gamma di Goodman e Kruskal
in tavole di contingenza, 372
gestione del rumore
in analisi Cluster TwoStep, 511
Get Data, 78
confronto con il comando Data List, 78
confronto con il comando Get Capture, 78
giornale di sessione, 630
giustificazione, 261, 632
output, 261, 632
GLM
grafici di profilo, 429
modello, 425
salvataggio di matrici, 433
salvataggio di variabili, 433
somma dei quadrati, 425
test Post Hoc, 431

GLM univariato, 421, 437
  contrasti, 428
  diagnostici, 436
  medie marginali stimate, 436
  opzioni, 436
  visualizzazione, 436
grafici, 259, 266–267, 270, 637, 662
  cenni generali, 605
  copia, 266
  copia in altre applicazioni, 266
  creazione, 605
  esportazione, 270, 662
  etichette dei casi, 473
  incollare in altre applicazioni, 267
  modelli, 616, 637
  modifica, 609
  nascondere, 259
  nella Curva ROC, 619
  note a piè di pagina, 613
  proporzioni, 637
  sottotitoli, 613
  titoli, 613
  valori mancanti, 614
grafici a barre
  in frequenze, 350
grafici a dispersione
  in regressione lineare, 463
grafici a scatole
  confronto dei livelli del fattore, 364
  confronto di variabili, 364
  in esplora, 364
grafici a stalattite
  in cluster gerarchica, 523
grafici a torta
  in frequenze, 350
  titoli, 613
grafici dei pesi fattoriali
  in analisi fattoriale, 501
grafici dei residui
  in GLM univariato, 436
grafici di importanza delle variabili
  in analisi Cluster TwoStep, 514
grafici di normalità detrendizzati
  in esplora, 364
grafici di probabilità normale
  in esplora, 364
  in regressione lineare, 463
grafici di profilo
  in GLM, 429
grafici di variabilità vs. densità
  in esplora, 364
  in GLM univariato, 436
grafici interattivi, 266–267, 641
  copia in altre applicazioni, 266
  incorporamento come oggetti ActiveX, 267
  opzioni, 641
  salvataggio di dati con un grafico, 641
grafici parziali
  in regressione lineare, 463
grafici ramo-foglia
  in esplora, 364
grafici: opzioni, 637
grafico di importanza
  in analisi Cluster TwoStep, 514
griglie, 322
  tabelle pivot, 322

H di Kruskal-Wallis
  in test per due campioni indipendenti, 554
Hochberg (GT2)
  in ANOVA univariata, 415
  in GLM, 431

HTML, 270, 274, 662, 747
aggiunta di codice personalizzato, 747
esportazione dell'output, 270, 274, 662

ICC. *Vedere* coefficiente di correlazione intraclasse, 589
icone, 656
modifica delle icone della barra degli strumenti, 656
nelle finestre di dialogo, 13
identificatori di comando, 717
impostazione della pagina, 285, 287–288
dimensioni del grafico, 288
intestazioni e piè di pagina, 287
incollare, 267–269
grafici, 267
oggetti speciali, 269
tabelle pivot, 267–268
tabelle pivot come tabelle, 268
indice di concentrazione
in Statistiche di rapporto, 603
informazioni sul file, 68–69
informazioni sulle variabili, 623
inserimento
grafici interattivi, 267
tabelle pivot, 267
inserimento di dati, 108–110
non numerici, 110
numerico, 109
utilizzo di etichette dei valori, 110
inserimento di etichette di gruppo, 304
inserimento di segnalibri nelle visualizzazioni delle tabelle pivot, 310
inserimento nella cache, 80
file attivo, 80
insiemi a risposta multipla
categorie multiple, 132
definizione, 132
dicotomie multiple, 132
Insiemi a risposta multipla, 562
categorie, 562
dicotomie, 562
etichette degli insiemi, 562
nomi degli insiemi, 562
insiemi di variabili, 625–626
definizione, 625
utilizzo, 626
interruzioni di riga
etichette di variabili e dei valori, 104
interruzioni di tabelle, 333
intervalli di confidenza
in ANOVA univariata, 418
in esplora, 363
in GLM, 428, 436
in regressione lineare, 468
in test T per campioni appaiati, 406
in test T per campioni indipendenti, 403
in test T per un campione, 408
nella Curva ROC, 622
salvataggio in regressione lineare, 465
intervalli di stima
salvataggio in regressione lineare, 465
salvataggio in stima di curve, 478
intervallo
in cubi OLAP, 394
in descrittive, 355
in frequenze, 348
in medie, 388
in Riassumi, 382
in Statistiche di rapporto, 603

intervallo multiplo di Ryan-Einot-Gabriel-Welsch
in ANOVA univariata, 415
in GLM, 431
intestazioni, 285, 287
istogrammi
in esplora, 364
in frequenze, 350
in regressione lineare, 463
iterazioni
in analisi fattoriale, 499, 501
nel cluster con metodo delle K-medie, 530

kappa
in tavole di contingenza, 372
kappa di Cohen
in tavole di contingenza, 372
KR20
in analisi di affidabilità, 589
Kuder-Richardson 20 (KR20)
in analisi di affidabilità, 589

LAG (funzione), 202
lambda
in tavole di contingenza, 372
lambda di Goodman e Kruskal
in tavole di contingenza, 372
Lambda di Wilks
in analisi discriminante, 486
larghezza di colonna, 106, 316, 324, 643
controllo della larghezza massima, 316
controllo della larghezza per il testo a capo, 316
controllo della larghezza predefinita, 643
nell'Editor dei dati, 106
tabelle pivot, 324
lingua
modifica della lingua dell'output, 630
linguaggio a comandi, 335
livellamento, 202
livellamento T4253H, 202
livello di misurazione, 99, 129
definizione, 99
icone nelle finestre di dialogo, 13
login al server, 85

margini, 285, 328
delle celle, 328
margini della pagina, 285
massima verosimiglianza
in analisi fattoriale, 499
massimo
confronto di colonne del report, 583
in cubi OLAP, 394
in descrittive, 355
in esplora, 363
in frequenze, 348
in medie, 388
in Riassumi, 382
in Statistiche di rapporto, 603
matrice di correlazione
in analisi discriminante, 485
in analisi fattoriale, 493, 498
matrice di covarianza
in analisi discriminante, 485, 488
in GLM, 433
in regressione lineare, 468
matrice di modelli
in analisi fattoriale, 493
matrice di trasformazione
in analisi fattoriale, 493
media
di più colonne del report, 583
in ANOVA univariata, 418

in cubi OLAP, 394
in descrittive, 355
in esplora, 363
in frequenze, 348
in medie, 388
in Report: Riepiloghi per colonne, 582
in Report: Riepiloghi per righe, 575
in Riassumi, 382
in Statistiche di rapporto, 603
sottogruppo, 385, 391
media armonica
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
media geometrica
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
media mobile centrata, 202
media mobile sui precedenti, 202
media pesata
in Statistiche di rapporto, 603
media trim
in esplora, 363
mediana
in cubi OLAP, 394
in esplora, 363
in frequenze, 348
in medie, 388
in Riassumi, 382
in Statistiche di rapporto, 603
mediana dei gruppi
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
mediana mobile, 202
Medie, 385, 676
di scala, 388
opzioni, 388
output ASCII da, 676
medie di gruppi, 385, 391
medie di sottogruppi, 385, 391
medie marginali stimate
in GLM univariato, 436
medie osservate
in GLM univariato, 436
memoria, 630
menu, 9, 651
personalizzazione, 651
metafile, 270, 277, 662
esportazione di grafici, 270, 277, 662
metodi, 694
oggetti di automazione OLE, 694
metodi di selezione, 331
selezione di righe e colonne nelle tabelle pivot, 331
Microsoft Access, 33
minimi quadrati generalizzati
in analisi fattoriale, 499
minimi quadrati non ponderati
in analisi fattoriale, 499
minimi quadrati ponderati
in regressione lineare, 457
minimo
confronto di colonne del report, 583
in cubi OLAP, 394
in descrittive, 355
in esplora, 363
in frequenze, 348
in medie, 388
in Riassumi, 382
in Statistiche di rapporto, 603

misura della differenza delle misure
in distanze, 453
misura delle differenze dei modelli
in distanze, 453
misura di dissimilarità di Lance e Williams, 453
in distanze, 453
misura di distanza phi-quadrato
in distanze, 453
misure di dispersione
in descrittive, 355
in esplora, 363
in frequenze, 348
in Statistiche di rapporto, 603
misure di distanza
in cluster gerarchica, 521
in distanze, 453
misure di distribuzione
in descrittive, 355
in frequenze, 348
misure di similarità
in cluster gerarchica, 521
in distanze, 454
misure di tendenza centrale
in esplora, 363
in frequenze, 348
in Statistiche di rapporto, 603
moda
in frequenze, 348
modalità distribuita, 83, 85, 88–90, 92, 667
accesso a file di dati, 88, 90
percorsi UNC, 92
procedure disponibili, 92
salvataggio di file di dati, 89
utilità SPSS Production, 667
modelli, 106, 108, 637
definizione di variabili, 106, 108
nei grafici, 616, 637
utilizzo di un file di dati esterno come modello, 136
modelli fattoriali completi
in GLM, 425
modelli personalizzati
in GLM, 425
modello composto
in stima di curve, 477
modello cubico
in stima di curve, 477
modello di crescita
in stima di curve, 477
modello di Guttman
in analisi di affidabilità, 587, 589
modello di potenza
in stima di curve, 477
modello esponenziale
in stima di curve, 477
modello inverso
in stima di curve, 477
modello lineare
in stima di curve, 477
modello logaritmico
in stima di curve, 477
modello logistico
in stima di curve, 477
modello parallelo
in analisi di affidabilità, 587, 589
modello parallelo esatto
in analisi di affidabilità, 587, 589
modello quadratico
in stima di curve, 477
modello S
in stima di curve, 477
modifica di dati, 111–112

moltiplicazione
  moltiplicazione tra colonne del report, 583
mostrare, 259, 312–313, 652
  barre degli strumenti, 652
  didascalie, 313
  etichette delle dimensioni, 312
  note a piè di pagina, 313
  righe o colonne, 312
  risultati, 259
  titoli, 313
mrInterview, 62

nascondere, 259–260, 312–313, 652
  barre degli strumenti, 652
  didascalie, 313
  etichette delle dimensioni, 312
  note a piè di pagina, 313
  righe e colonne, 312
  risultati della procedura, 260
  titoli, 313
nascondere (escludere) l'output da Viewer con SGO, 726
Newman-Keuls
  in GLM, 431
nomi di variabili, 99, 630
  file in formato portabile, 74
  generati da SGO, 735
  nelle finestre di dialogo, 11, 630
  nomi di variabili a casi misti, 99
  nomi lunghi di variabili con testo a capo nell'output, 99
  regole, 99
  troncamento dei nomi di variabili lunghe nelle versioni precedenti, 74
nominale, 99
  livello di misurazione, 99, 129
notazione scientifica, 100, 326, 630
  nelle tabelle pivot, 326
  soppressione nell'output, 630
note a piè di pagina, 317, 330, 332
  aggiunta a una tabella, 332
  nei grafici, 613
  rinumerazione, 330
  simboli, 317, 330
novità
  SPSS 14.0, 2
numerazione delle pagine, 285, 288
  nei report di riepilogo per colonne, 585
  nei report di riepilogo per righe, 577
numeri di porta, 85
numero di casi
  in cubi OLAP, 394
  in medie, 388
  in Riassumi, 382

oggetti, 689, 691–692, 695
  cenni generali, 689, 691
  convenzioni di denominazione delle variabili, 691
  utilizzo negli script, 689, 692, 695
oggetti ActiveX, 267
oggetti di automazione, 689, 691–692, 694–695
  convenzioni di denominazione delle variabili, 691
  metodi, 694
  proprietà, 694
  tipi, 691
  utilizzo negli script, 689, 692, 695
  Visualizzatore oggetti, 695
ombreggiatura, 329
  delle celle, 329

OMS
  utilizzo di XSLT con OXML, 742
opzioni, 630, 632–633, 635, 637, 641, 643, 645–646, 648
  anni a due cifre, 645
  aspetto della tabella pivot, 643
  Bozza, 633
  dati, 645
  directory temporanea, 630
  etichette di output, 635
  generali, 630
  grafici, 637
  grafici interattivi, 641
  script, 648
  valuta, 646
  Viewer, 632
opzioni della riga di comando, 671
  utilità SPSS Production, 671
ordinale, 99
  livello di misurazione, 99, 129
ordinamento dei casi, 212
ordine di visualizzazione, 303
output, 257, 259–261, 265–266, 269–270, 289, 291, 331, 632, 662
  allineamento, 261, 632
  bozza, 291
  centratura, 261, 632
  copia, 260–261
  copia in altre applicazioni, 266
  copiare e incollare più elementi, 269
  eliminazione, 260
  esportazione, 270, 662
  incollare in altre applicazioni, 265
  modifica, 331
  modifica della lingua dell'output, 630
  mostrare, 259
  nascondere, 259
  salvare, 289
  spostamento, 260
  Viewer, 257
OXML, 742

percentili
  in esplora, 363
  in frequenze, 348
percentuali
  in tavole di contingenza, 376
percentuali complessive
  in tavole di contingenza, 376
percentuali di colonna
  in tavole di contingenza, 376
percentuali di riga
  in tavole di contingenza, 376
permessi, 92
phi
  in tavole di contingenza, 372
piè di pagina, 285, 287
pivoting
  controllo con l'output esportato di SGO, 733
ponderazione di casi, 232
  pesi frazionari nelle tavole di contingenza, 232
PowerPoint, 274
  esportazione dell'output come PowerPoint, 274
prestazioni, 80
  copia di dati nella cache, 80
previsione
  in stima di curve, 478
primo
  in cubi OLAP, 394
  in medie, 388
  in Riassumi, 382

procedure, 696
  script, 696
procedure globali, 648, 697
produzione automatica, 659
profondità struttura
  in analisi Cluster TwoStep, 511
programmazione con il linguaggio a comandi, 335
proporzioni, 637
proprietà, 316, 694
  oggetti di automazione OLE, 694
  sovrapposte, 316
  tabelle pivot, 316
protezione tramite password, 290
pubblicazione dell'output, 674
  con l'utilità SPSS Production, 673
pulsante ?, 12
Pulsante di annullamento, 12
Pulsante di ripristino, 12
pulsante Incolla, 12
Pulsante OK, 12
pulsanti, 656
  modifica delle icone della barra degli strumenti, 656
punteggi esponenziali, 177
punteggi fattoriali, 502
punteggi fattoriali di Anderson-Rubin, 502
punteggi fattoriali di Bartlett, 502
punteggi normali
  in classifica casi, 177
punteggi *z*
  in classifica casi, 177
  in descrittive, 353
  salvataggio come variabili, 353
punteggio, 205
  caricamento di modelli salvati, 206
  visualizzazione modelli caricati, 210
punti di interruzione, 704
  negli script, 704

Q di Cochran
  in test per diversi campioni dipendenti, 559
Quancept, 62
Quanvert, 62
quartili
  in frequenze, 348

R *multiplo*
  in regressione lineare, 468
R-E-G-W *F*
  in ANOVA univariata, 415
  in GLM, 431
R-E-G-W *Q*
  in ANOVA univariata, 415
  in GLM, 431
R^2
  in medie, 388
  in regressione lineare, 468
  modifica di R^2, 468
R^2 corretto
  in regressione lineare, 468
raggruppamento
  scelta di una procedura, 505
raggruppamento di righe o colonne, 304
ramificazioni massime
  in analisi Cluster TwoStep, 511
rapporto di covarianza
  in regressione lineare, 465
regressione
  grafici, 463
  Regressione lineare, 457
  regressione multipla, 457

Regressione lineare, 457
  blocchi, 457
  di scala, 468
  esportazione di informazioni dei modelli, 465
  grafici, 463
  metodi di selezione delle variabili, 461, 470
  opzioni aggiuntive del comando, 471
  pesi, 457
  residui, 465
  salvataggio di nuove variabili, 465
  valori mancanti, 470
  variabile di selezione, 463
regressione multipla
  in regressione lineare, 457
report
  colonne totale, 583
  confronto di colonne, 583
  divisione di valori di colonne, 583
  moltiplicazione di valori di colonne, 583
  report di riepilogo per colonne, 580
  report di riepilogo per righe, 571
  totali composti, 583
report di riepilogo per colonne, 580
Report: Riepiloghi per colonne, 580
  colonne totale, 583
  controllo pagina, 584
  formato colonne, 574
  layout di pagina, 577
  numerazione delle pagine, 585
  opzioni aggiuntive del comando, 586
  totale finale, 585
  totali parziali, 584
  valori mancanti, 585
Report: Riepiloghi per righe, 571
  colonne di dati, 571
  colonne di separazione, 571
  controllo pagina, 576
  formato colonne, 574
  layout di pagina, 577
  numerazione delle pagine, 577
  opzioni aggiuntive del comando, 586
  ordinamento delle sequenze, 571
  piè di pagina, 579
  spaziatura separazioni, 576
  titoli, 579
  valori mancanti, 577
  variabili nei titoli, 579
residui
  in tavole di contingenza, 376
  salvataggio in regressione lineare, 465
  salvataggio in stima di curve, 478
residui cancellati
  in GLM, 433
  in regressione lineare, 465
residui non standardizzati
  in GLM, 433
residui standardizzati
  in GLM, 433
  in regressione lineare, 465
residui studentizzati
  in regressione lineare, 465
rho
  in correlazioni bivariate, 439
  in tavole di contingenza, 372
Riassumi, 379
  di scala, 382
  opzioni, 381
ricodifica di valori, 149, 169–170, 172–173, 179
ridenominazione dei file di dati, 122
ridimensionamento delle tabelle pivot, 316, 322
righe, 331
  selezione nelle tabelle pivot, 331

rimozione di etichette di gruppo, 304
riordinamento di righe e colonne, 303
rischio
in tavole di contingenza, 372
Risposte multiple
opzioni aggiuntive del comando, 570
Risposte multiple: Frequenze, 564
valori mancanti, 564
Risposte multiple: Tavole di contingenza, 566
associazione di variabili negli insiemi a risposta multipla, 569
definizione degli intervalli dei valori, 568
percentuali basate sui casi, 569
percentuali basate sulle risposte, 569
percentuali nelle celle, 569
valori mancanti, 569
ristrutturazione dei dati, 234, 237–240, 243–245, 247–248, 250–251, 253, 255
cenni generali, 234
creazione di più variabili indice per variabili in casi, 247
creazione di una variabile indice per variabili in casi, 245
creazione di variabili indice per variabili in casi, 243
e dati pesati, 255
esempio di casi in variabili, 238
esempio di due indici per variabili in casi, 245
esempio di indice singolo per variabili in casi, 244
esempio di variabili in casi, 237
gruppi di variabili per variabili in casi, 239
opzioni per casi in variabili, 253
opzioni per variabili in casi, 248
ordinamento di dati per casi in variabili, 251
selezione di dati correnti per variabili in casi, 240
selezione di dati per casi in variabili, 250
tipi di ristrutturazione, 234
rotazione di etichette, 305
rotazione equamax
in analisi fattoriale, 501
rotazione oblimin diretta
in analisi fattoriale, 501
rotazione quartimax
in analisi fattoriale, 501
rotazione varimax
in analisi fattoriale, 501

s-stress
in scaling multidimensionale, 593
salvataggio dell'output, 270, 274–275, 290, 300, 662, 674
formato Excel, 270, 274
Formato PowerPoint, 270, 274
formato testo, 270, 275, 662
formato Word, 270, 274
HTML, 270, 274, 662
output bozza, 300
protezione tramite password, 290
pubblicazione sul Web, 674
salvataggio dell'output della Bozza come testo, 300
salvataggio di file, 69, 73–74, 89
file di dati, 73–74, 89
file di dati SPSS, 69
query di file di database, 49
salvataggio di grafici, 270, 277–281, 641, 662
file BMP, 270, 277–278, 662
file EPS, 270, 277, 280, 662
file JPEG, 270, 277–278, 662
file PICT, 270, 277–278, 662
file PNG, 279

file PostScript, 280
file TIFF, 279
file WMF, 270, 277, 281
formato WMF, 662
metafile, 270, 277, 662
salvataggio di grafici interattivi con dati, 641
salvataggio di visualizzazioni delle tabelle pivot, 310
scala, 99
in analisi di affidabilità, 587
in scaling multidimensionale, 593
livello di misurazione, 99, 129
scaling di grafici esportati, 277
Scaling multidimensionale, 593
condizionalità, 597
creazione di matrici delle distanze, 596
criteri, 598
definizione della forma dei dati, 595
di scala, 593
dimensioni, 597
esempio, 593
livelli di misurazione, 597
misure di distanza, 596
modelli di scaling, 597
opzioni aggiuntive del comando, 599
opzioni di visualizzazione, 598
trasformazione di valori, 596
Scheda Immediate, 705
finestra dello script, 705
Scheda Loaded, 705
finestra dello script, 705
Scheda Stack, 705
finestra dello script, 705
Scheda Watch, 705
finestra dello script, 705
scomposizione gerarchica, 426
script, 648, 651, 656, 675–677, 680, 683–684, 699, 704
aggiungere una descrizione, 699
aggiunta a menu, 651
cenni generali, 675
con sintassi dei comandi, 707, 709
creazione, 680, 684
debug, 704–705
dichiarazione di variabili, 690–691
esecuzione, 675
esecuzione con i pulsanti della barra degli strumenti, 656
file degli script automatici, 648, 688
file delle procedure globali, 648, 697
finestra dello script, 680, 682
finestre di dialogo, 699, 701
script automatici, 677, 684, 688
script di avvio, 683
utilizzo di oggetti di automazione, 689, 691–692, 695
script automatici, 648, 684, 686
creazione, 684
eventi di attivazione, 686
file degli script automatici, 688
script di avvio, 683
script globali, 697
segmentazione di valori dei dati, 149
segnalibri, 310
segno di dollaro, 326
nelle tabelle pivot, 326
segno di percentuale, 326
nelle tabelle pivot, 326
selezione di casi, 227
campione casuale, 230
in base a criteri di selezione, 229
intervallo di casi, 231

intervallo di dati, 231
intervallo di orari, 231
selezione in avanti
in regressione lineare, 461
selezione per passi
in regressione lineare, 461
seme dei numeri casuali, 165
server, 85
aggiunta, 85
login, 85
modifica, 85
nomi, 85
numeri di porta, 85
server remoti, 83, 85, 88–90, 92, 667
accesso a file di dati, 88, 90
aggiunta, 85
login, 85
modifica, 85
percorsi UNC, 92
procedure disponibili, 92
salvataggio di file di dati, 89
utilità SPSS Production, 667
SGO, 711, 741
controllo delle tabelle pivot, 720, 733
esclusione dell'output da Viewer, 726
formato file dati SPSS, 720, 726
formato file sav, 720, 726
formato testo, 720
identificatori di comando, 717
nomi di variabili nei file SAV, 735
sottotipi di tabelle, 717
tipi di oggetti di output, 715
XML, 720, 737
sintassi, 335, 342, 630, 632–633, 655, 659, 707, 709
accesso a *SPSS Command Syntax Reference*, 17
con script, 707, 709
esecuzione, 342
esecuzione di sintassi dei comandi con i pulsanti della barra degli strumenti, 655
file giornale, 340, 342, 630
incollare, 337
incollare negli script, 708
registro, 632–633
registro di output, 339
regole dell'utilità SPSS Production, 659
regole sintassi, 335
sintassi dei comandi, 335, 342, 630, 632–633, 651, 656, 659
accesso a *SPSS Command Syntax Reference*, 17
aggiunta a menu, 651
esecuzione, 342
esecuzione con i pulsanti della barra degli strumenti, 656
file giornale, 340, 342, 630
formattazione dell'utilità SPSS Production, 670
incollare, 337
registro, 632–633
registro di output, 339
regole dell'utilità SPSS Production, 659
regole sintassi, 335
Sistema di gestione dell'output (SGO), 711, 741
sistema di misurazione, 630
soglia iniziale
in analisi Cluster TwoStep, 511
somma
in cubi OLAP, 394
in descrittive, 355
in frequenze, 348
in medie, 388
in Riassumi, 382
somma cumulata, 202

somma dei quadrati, 426
in GLM, 425
sostituzione di valori mancanti
interpolazione lineare, 205
media della serie., 205
media di punti vicini., 205
mediana di punti vicini, 205
trend lineare, 205
sottoinsiemi di casi
campione casuale, 230
selezione, 227, 229, 231
sottotipi, 717
vs. etichette, 718
sottotipi di tabelle, 717
vs. etichette, 718
sottotitoli
nei grafici, 613
sovrapposte, 13, 98, 113–114, 234, 333, 623, 625, 630
controllo delle interruzioni di tabella, 333
definizione, 98
definizione di insiemi di variabili, 625
informazioni sulla definizione, 623
informazioni sulle variabili nelle finestre di dialogo, 13
inserimento di nuove variabili, 113
nelle finestre di dialogo, 10
ordine di visualizzazione nelle finestre di dialogo, 630
ricodifica, 169–170, 172–173, 179
ridenominazione di file di dati uniti, 217
ristrutturazione in casi, 234
selezione nelle finestre di dialogo, 13
spostamento, 114
spazio su disco, 78, 80
temporaneo, 78, 80
spazio su disco temporaneo, 78, 80
spostamento di righe e colonne, 303
SPSS
passaggi fondamentali, 14
stampa, 117, 282–283, 285, 287–288, 298–299, 316, 322, 333
anteprima di stampa, 283
controllo delle interruzioni di tabella, 333
dati, 117
dimensioni del grafico, 288
grafici, 282
impostazione della pagina, 285
intestazioni e piè di pagina, 285, 287
numeri di pagina, 288
output bozza, 298–299
output testuale, 282
ridimensionamento di tabelle, 316, 322
spaziatura tra elementi dell'output, 288
strati, 282, 316, 322
tabelle pivot, 282
standardizzazione
in analisi Cluster TwoStep, 511
standardizzazione di valori
in descrittive, 353
statistica di Brown-Forsythe
in ANOVA univariata, 418
statistica di Cochran
in tavole di contingenza, 372
statistica di Mantel-Haenszel
in tavole di contingenza, 372
statistica di Welch
in ANOVA univariata, 418
statistiche descrittive
in analisi Cluster TwoStep, 515
in descrittive, 353
in esplora, 363

in frequenze, 348
in GLM univariato, 436
in Riassumi, 382
in Statistiche di rapporto, 603
Statistiche di rapporto, 601
di scala, 603
statistiche Durbin-Watson
in regressione lineare, 468
statistiche *R*
in medie, 388
in regressione lineare, 468
Stima di curve, 473
analisi della varianza, 473
inclusione di costanti, 473
modelli, 477
previsione, 478
salvataggio degli intervalli di stima, 478
salvataggio dei residui, 478
salvataggio di valori attesi, 478
stimatore di Tuckey a doppio peso
in esplora, 363
stimatore M di Andrew
in esplora, 363
Stimatore M di Hampel
in esplora, 363
stimatore M di Huber
in esplora, 363
stimatori M
in esplora, 363
stime dei parametri
in GLM univariato, 436
stime di Blom, 177
stime di proporzione
in classifica casi, 177
stime di Rankit, 177
stime di Tukey, 177
stime di Van der Waerden, 177
stime effetto-dimensioni
in GLM univariato, 436
stime potenza
in GLM univariato, 436
strati, 282, 306, 308–309, 316, 322
creazione, 306
in tavole di contingenza, 372
modifica, 308
nelle tabelle pivot, 306
stampa, 282, 316, 322
visualizzazione, 306, 309
stress
in scaling multidimensionale, 593
Student-Newman-Keuls
in ANOVA univariata, 415
in GLM, 431
studio di casi di controllo
Test T per campioni appaiati, 403
studio di confronti tra coppie
in test T per campioni appaiati, 403
suddivisione di tabelle, 333
controllo delle interruzioni di tabella, 333
suggerimenti per gli script, 675, 680, 683, 689–690, 692, 694–696, 699, 704
aggiungere una descrizione, 699
debug, 704
dichiarazioni di variabili, 690
finestra dello script, 680
finestre di dialogo personalizzate, 699
modalità di funzionamento degli script, 689
ottenere oggetti di automazione, 692
procedure, 696
proprietà e metodi, 694
script di avvio, 683
Visualizzatore oggetti, 695

tabelle di frequenza
in esplora, 363
in frequenze, 345
tabelle pivot, 259, 265–270, 282, 301–306, 311, 314, 316–317, 319–320, 322, 324, 331–333, 643, 662, 668
adattamento alla pagina, 316, 322
adattamento predefinito della larghezza della colonna, 643
aggiunta di didascalie, 332
aspetto predefinito per le nuove tabelle, 643
bordi, 320
controllo del formato per le sessioni dell'utilità SPSS Production, 668
controllo delle interruzioni di tabella, 333
copia, 266
copia in altre applicazioni, 266
copiare e incollare più tabelle, 269
eliminazione di etichette di gruppo, 304
esportazione come HTML, 270, 662
formati di celle, 319
gestione, 301
griglie, 322
identificazione di dimensioni, 302
incollare come metafile, 267
incollare come tabelle, 265, 268
incollare come testo, 268
incollare in altre applicazioni, 265
incorporamento come oggetti ActiveX, 267
inserimento di etichette di gruppo, 304
larghezza delle celle, 324
modifica, 301
modifica dell'aspetto, 314
modifica dell'ordine di visualizzazione, 303
modifica di due o più, 301
nascondere, 259
pivoting, 301–302
proprietà, 316
proprietà delle note a piè di pagina, 317
proprietà generali, 316
raggruppamento di righe o colonne, 304
reimpostazione dei valori predefiniti, 305
ricerca di definizioni di etichetta, 306
rotazione di etichette, 305
selezione di righe e colonne, 331
separazione di righe o colonne, 304
spostamento di righe e colonne, 303
stampa di strati, 282
stampa di tabelle di grandi dimensioni, 333
strati, 306
testo di prosecuzione, 322
trasposizione di righe e colonne, 303
utilizzo di icone, 302
visualizzazione dei bordi nascosti, 322
visualizzazione delle celle, 311
TableLook, 314–315
applicazione, 314
creazione, 315
Tamhane (T2)
in ANOVA univariata, 415
in GLM, 431
tau di Goodman e Kruskal
in tavole di contingenza, 372
tau di Kruskal
in tavole di contingenza, 372
tau-*b*
in tavole di contingenza, 372
Tau-*b* di Kendall
in correlazioni bivariate, 439
in tavole di contingenza, 372
tau-*c*
in tavole di contingenza, 372

tau-*c* di Kendall , 372
in tavole di contingenza, 372
tavole di contingenza, 369
in tavole di contingenza, 369
risposte multiple, 566
Tavole di contingenza, 369
di scala, 372
formati, 377
grafici a barre raggruppati, 372
pesi frazionari, 232
soppressione di tabelle, 369
strati, 372
variabili di controllo, 372
visualizzazione cella, 376
termini di interazione, 426
test a intervallo multiplo di Duncan
in ANOVA univariata, 415
in GLM, 431
Test binomiale, 538
di scala, 540
dicotomie, 538
opzioni, 540
opzioni aggiuntive del comando, 540
valori mancanti, 540
test del segno
in test per due campioni dipendenti, 551
test della mediana
in test per due campioni indipendenti, 554
Test delle reazioni estreme di Moses
in test per due campioni indipendenti, 549
Test delle successioni, 541
di scala, 543
opzioni, 543
opzioni aggiuntive del comando, 543
punti di divisione, 541–542
valori mancanti, 543
Test delle successioni di Wald-Wolfowitz
in test per due campioni indipendenti, 549
test di additività di Tukey
in analisi di affidabilità, 587, 589
Test *di* Dunnett
in ANOVA univariata, 415
in GLM, 431
Test di Kolmogorov-Smirnov per un campione, 544
di scala, 546
distribuzione di test, 544
opzioni, 546
opzioni aggiuntive del comando, 546
valori mancanti, 546
test di Levene
in ANOVA univariata, 418
in esplora, 364
in GLM univariato, 436
test di Lilliefors
in esplora, 364
test di linearità
in medie, 388
test di McNemar
in tavole di contingenza, 372
in test per due campioni dipendenti, 551
test di normalità
in esplora, 364
test di omogeneità della varianza
in ANOVA univariata, 418
in GLM univariato, 436
test di Scheffé
in ANOVA univariata, 415
in GLM, 431
test di sfericità di Bartlett
in analisi fattoriale, 498
test di Shapiro-Wilk
in esplora, 364

Test di Sidak
in ANOVA univariata, 415
in GLM, 431
Test *di* Tukey
in ANOVA univariata, 415
in GLM, 431
Test *di* Waller-Duncan
in ANOVA univariata, 415
in GLM, 431
test di Wilcoxon
in test per due campioni dipendenti, 551
test esatto di Fisher
in tavole di contingenza, 372
test M di Box
in analisi discriminante, 485
test non parametrici
chi-quadrato, 534
Test delle successioni, 541
Test di Kolmogorov-Smirnov per un campione, 544
Test per diversi campioni dipendenti, 558
Test per diversi campioni indipendenti, 554
Test per due campioni dipendenti, 551
Test per due campioni indipendenti, 547
Test per confronti a coppie di Gabriel
in ANOVA univariata, 415
in GLM, 431
Test per confronti a coppie di Games e Howell
in ANOVA univariata, 415
in GLM, 431
Test per diversi campioni dipendenti, 558
di scala, 560
opzioni aggiuntive del comando, 560
tipi di test, 559
Test per diversi campioni indipendenti, 554
definizione dell'intervallo, 557
di scala, 557
opzioni, 557
opzioni aggiuntive del comando, 558
tipi di test, 556
valori mancanti, 557
variabili di raggruppamento, 557
Test per due campioni dipendenti, 551
di scala, 554
opzioni, 554
opzioni aggiuntive del comando, 554
tipi di test, 553
valori mancanti, 554
Test per due campioni indipendenti, 547
definizione dei gruppi, 550
di scala, 550
opzioni, 550
opzioni aggiuntive del comando, 551
tipi di test, 549
valori mancanti, 550
variabili di raggruppamento, 550
test per l'indipendenza
chi-quadrato, 372
test *t*
in GLM univariato, 436
in test T per campioni appaiati, 403
in test T per campioni indipendenti, 399
in test T per un campione, 406
Test *T* di Student, 399
Test T per campioni appaiati, 403
opzioni, 406
selezione di variabili appaiate, 403
valori mancanti, 406
test T per campioni indipendenti, 399
definizione dei gruppi, 402
intervalli di confidenza, 403
opzioni, 403

valori mancanti, 403
variabili di raggruppamento, 402
variabili stringa, 402
Test *T* per due campioni
in test T per campioni indipendenti, 399
Test T per un campione, 406
intervalli di confidenza, 408
opzioni, 408
opzioni aggiuntive del comando, 409
valori mancanti, 408
testi di Friedman
in test per diversi campioni dipendenti, 559
testo, 51, 264–265, 270, 275, 291, 300, 331, 662
aggiunta al Viewer, 264
aggiunta di un file di testo al Viewer, 265
creazione di output testuale, 291
delle celle, 331
esportazione dell'output come testo, 270, 275, 300, 662
file di dati, 51
testo a capo, 316
controllo della larghezza della colonna per il testo a capo, 316
etichette di variabili e dei valori, 104
testo di etichetta verticale, 305
testo di prosecuzione, 322
per tabelle pivot, 322
test*t* dipendenti
in test T per campioni appaiati, 403
tipi di dati, 100, 102, 114, 646
definizione, 100
formati di input, 102
formati di visualizzazione, 102
modifica, 114
valuta personalizzata, 100, 646
tipi di oggetti di output
in SGO, 715
titoli, 264
aggiunta al Viewer, 264
in cubi OLAP, 398
nei grafici, 613
tolleranza
in regressione lineare, 468
totali, 676
formattazione in grassetto automatica nell'output, 676
totali finali
nei report di riepilogo per colonne, 585
totali parziali
nei report di riepilogo per colonne, 584
$T^{quadrato}$ di Hotelling
in analisi di affidabilità, 587, 589
trasformazioni di dati, 645
calcolo di variabili, 161
classifica di casi, 175
condizioni, 163
funzioni, 164
ricodifica di valori, 169–170, 172–173, 179
ritardo di esecuzione, 645
serie storiche, 198, 201
variabili stringa, 164
trasformazioni di file, 234
aggregazione dei dati, 221
distinzione per gruppi, 226
ordinamento dei casi, 212
ponderazione di casi, 232
ristrutturazione dei dati, 234
trasposizione di casi e variabili., 213
unione di file di dati, 214, 218
trasposizione di casi e variabili., 213
trasposizione di righe e colonne, 303

U di Mann-Whitney
in test per due campioni indipendenti, 549
ultimo
in cubi OLAP, 394
in medie, 388
in Riassumi, 382
unione di file di dati
file con casi diversi, 214
file con variabili diverse, 218
informazioni del dizionario, 218
ridenominazione di variabili, 217
unità condivise, 92
utilità SPSS Production, 630, 632–633, 659, 662, 664, 666–668, 671, 673
controllo del formato con la sintassi dei comandi, 670
controllo del formato per le tabelle pivot, 668
esecuzione di più lavori di produzione, 671
esportazione dell'output, 662
esportazione di grafici, 659, 662
file di output, 659
opzioni, 667
opzioni della riga di comando, 671
pianificazione di sessioni dell'utilità SPSS Production, 671
prompt macro, 666
pubblicazione dell'output, 673
pubblicazione sul Web, 673
regole sintassi, 659
sostituzione dei valori nei file di sintassi, 664
specifica di un server remoto, 667
utilizzo della sintassi dei comandi dal file giornale, 630
utilizzo della sintassi dei comandi dal registro, 632–633

V
in tavole di contingenza, 372
V di Rao
in analisi discriminante, 486
valori, 326
formato di visualizzazione delle tabelle pivot, 326
valori anomali
in analisi Cluster TwoStep, 511
in esplora, 363
in regressione lineare, 463
valori attesi
salvataggio in regressione lineare, 465
salvataggio in stima di curve, 478
valori attesi ponderati
in GLM, 433
valori d'influenza
in GLM, 433
in regressione lineare, 465
valori estremi
in esplora, 363
valori mancanti, 105
definizione, 105
in analisi fattoriale, 503
in ANOVA univariata, 418
in correlazioni bivariate, 442
in Correlazioni parziali, 448
in esplora, 366
in regressione lineare, 470
in Report: Riepiloghi per righe, 577
in Risposte multiple: Frequenze, 564
in Risposte multiple: Tavole di contingenza, 569
in test di Kolmogorov-Smirnov per un campione, 546
in test per due campioni dipendenti, 554
in test per due campioni indipendenti, 550

in test T per campioni appaiati, 406
in test T per campioni indipendenti, 403
in test T per un campione, 408
nei grafici, 614
nei report di riepilogo per colonne, 585
nei test per diversi campioni indipendenti, 557
nel test binomiale, 540
nel test chi-quadrato, 537
nel Test delle successioni, 543
nella Curva ROC, 622
nelle funzioni, 165
sostituzione in dati di serie storiche, 203
variabili stringa, 105
valori mancanti definiti dall'utente, 105
variabile di ambiente SPSSTMPDIR, 630
variabile di selezione
in regressione lineare, 463
variabili data
aggiungi o sottrai da variabili data/ora, 183
crea variabile data/ora da un insieme di variabili, 183
crea variabile data/ora dalla stringa, 183
definizione di dati di serie storiche, 199
estrai parte della variabile data/ora, 183
variabili di ambiente, 630
SPSSTMPDIR, 630
variabili di controllo
in tavole di contingenza, 372
variabili di raggruppamento, 234
creazione, 234
variabili di scala
segmentazione per la creazione di variabili categoriali, 149
variabili di separazione
in Aggrega dati, 221
variabili stringa, 105, 110
calcolo di nuove variabili stringa, 164
divisione delle stringhe lunghe nelle versioni precedenti, 74
inserimento di dati, 110
nelle finestre di dialogo, 10
ricodifica in interi consecutivi, 179
valori mancanti, 105
varianza
in cubi OLAP, 394
in descrittive, 355
in esplora, 363
in frequenze, 348
in medie, 388
in Report: Riepiloghi per colonne, 582
in Report: Riepiloghi per righe, 575
in Riassumi, 382
velocità, 80
copia di dati nella cache, 80
Viewer, 257, 259–264, 269, 288–289, 632, 635
compressione della vista riassuntiva, 263
copia dell'output, 261
eliminazione dell'output, 260
esclusione dei tipi di output con SGO, 726
espansione della vista riassuntiva, 263
finestra dei risultati, 257
incollare oggetti speciali, 269
modifica dei livelli della vista riassuntiva, 263
modifica del carattere della vista riassuntiva, 264
modifica delle dimensioni della vista riassuntiva, 264
nascondere risultati, 259
opzioni di visualizzazione, 632
salvataggio di documenti, 289
spaziatura tra elementi dell'output, 288
spostamento dell'output, 260

vista riassuntiva, 257, 262
visualizzazione di etichette dei valori, 635
visualizzazione di etichette di variabili, 635
visualizzazione di nomi di variabili, 635
visualizzazione di valori di dati, 635
vista riassuntiva, 262–263
compressione, 263
espansione, 263
modifica dei livelli, 263
nel Viewer, 262
viste/riquadri multipli
Editor dei dati, 115
Visualizzatore oggetti, 695
Visualizzazione dati, 96
Visualizzazione variabili, 97

W di Kendall
in test per diversi campioni dipendenti, 559
Web, 674
pubblicazione dell'output, 674

XML
output OXML da SGO, 742
salvataggio dell'output come XML, 711
struttura della tabella in OXML, 737
trasmissione dell'output a XML, 720
XSLT
utilizzo con OXML, 742

Z di Kolmogorov-Smirnov
in test di Kolmogorov-Smirnov per un campione, 544
in test per due campioni indipendenti, 549